

A Subfamily of Putative Cytokinin Receptors Is Revealed by an Analysis of the Evolution of the Two-Component Signaling System of Plants^{1[W][OPEN]}

Nijuscha Gruhn, Mhyeddeen Halawa, Berend Snel, Michael F. Seidl², and Alexander Heyl*

Institute of Biology/Applied Genetics, Dahlem Centre of Plant Sciences, Freie Universität, 14195 Berlin, Germany (N.G., M.H., A.H.); Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, 3584 CH, Utrecht, The Netherlands (B.S., M.F.S.); and Centre for BioSystems Genomics, 6700 AB, Wageningen, The Netherlands (B.S., M.F.S.)

The two-component signaling system—the major signaling pathway of bacteria—is found among higher eukaryotes only in plants, where it regulates diverse processes, such as the signaling of the phytohormone cytokinin. Cytokinin is perceived by a hybrid histidine (His) kinase receptor, and the signal is transduced by a multistep phosphorelay system of His phosphotransfer proteins and different classes of response regulators (RRs). To shed light on the origin and evolution of the two-component signaling system members in plants, we conducted a comprehensive domain-based phylogenetic study across the relevant kingdoms, including Charophyceae algae, the group of green algae giving rise to land plants. Surprisingly, we identified a subfamily of cytokinin receptors with members only from the early diverging land plants *Marchantia polymorpha* and *Physcomitrella patens* and then experimentally characterized two members of this subfamily. His phosphotransfer proteins of Charophyceae seemed to be more closely related to land plants than to other groups of green algae. Farther down the signaling pathway, the type-B RRs were found across all plant clades, but many members lack either the canonical Asp residue or the DNA binding domain. In contrast, the type-A RRs seemed to be limited to land plants. Finally, the analysis provided hints that one additional group of RRs, the type-C RRs, might be degenerated receptors and thus, of a different evolutionary origin than bona fide RRs.

Starting out as unicellular algae, plants have undergone many dramatic changes, enabling them to make major modifications in lifestyle, such as the transition from a single cell to multicellularity or from an aquatic to a terrestrial habitat (Rensing et al., 2008; Prochnik et al., 2010; Cock and Coelho, 2011). Implicit in these adaptations is the evolution of complex developmental programs. The execution of those programs is regulated by a multilayered interplay of different plant hormones (Jaillais and Chory, 2010; Vanstraelen and Benková, 2012; El-Showk et al., 2013).

One class of phytohormones is a group of *N*⁶-substituted adenine derivatives, the cytokinins. They have

been shown to act as plant growth regulators, crucial for plant development and for the response of plants to biotic and abiotic stress (Argueso et al., 2009; Werner and Schmülling, 2009; Choi et al., 2011; Brenner et al., 2012; Ha et al., 2012; Hwang et al., 2012). The cytokinin signal transduction is based on a variation of a signaling system common among bacteria, the two-component signaling (TCS) system. However, bacteria do not respond to cytokinin (Spichal, 2012). In its simplest form, the TCS system consists of a receptor His kinase, which auto-phosphorylates on signal perception, and a response regulator (RR), which mediates the output after being activated by phosphorylation of a canonical Asp in the RR domain. Cytokinin receptors are hybrid His kinases, because they contain both an His kinase and an RR domain. The cytokinin ligand is bound through the cyclase/His kinase-associated sensory extracellular (CHASE) domain (Anantharaman and Aravind, 2001; Mougél and Zhulin, 2001; Heyl et al., 2007), and this binding is thought to trigger a conformational change, leading to the auto-phosphorylation of the receptor (Miwa et al., 2007; Hothorn et al., 2011). After an intramolecular transfer from the His kinase to the RR domain of the receptor, the phosphate is transferred to His phosphotransfer (HPTs) proteins. These proteins shuttle continuously between the cytoplasm and the nucleus (Punwani et al., 2010). In the nucleus, they can phosphorylate the type-B RRs (RRBs). The RRBs are transcription factors containing an RR domain and an myeloblastosis (Myb)-related DNA binding domain, which allows them to bind to their target DNA

¹ This work was supported by the Volkswagen Foundation (grant no. Az I/83 477 to N.G.), Damascus University (grant no. 278), and the Dahlem Centre of Plant Sciences (to M.H.); parts of this project were carried out within the research program of the Centre of BioSystems Genomics, which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

² Present address: Laboratory of Phytopathology, Wageningen University, 6700 EE, Wageningen, The Netherlands.

* Address correspondence to hey1@zedat.fu-berlin.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Alexander Heyl (hey1@zedat.fu-berlin.de).

^[W] The online version of this article contains Web-only data.

^[OPEN] Articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.113.228080

sequences (Sakai et al., 2000; Hosoda et al., 2002). One group of their target genes is the type-A RRs (RRAs), which have been shown to work as negative regulators of the cytokinin signal transduction pathway (Hwang and Sheen, 2001; To et al., 2004). Most of the research on the cytokinin regulatory system has been carried out in the model plant *Arabidopsis thaliana* and to a lesser extent, also in other plants (Hellmann et al., 2010). Previous studies identified additional groups of RRs with the pseudo-RRs (PRRs), which have members that were shown to have a role in the regulation of circadian rhythm, and the type-C RRs (RRCs), for which a clear biological function has yet to be determined (Mizuno, 2004; Horák et al., 2008).

The ability of plants to use cytokinin as a phytohormone represents an evolutionary novelty (Gruhn and Heyl, 2013), which raises the questions of how a group of ubiquitous adenine derivatives became specifically regulated signaling molecules and how the required regulatory system, known from modern land plants, evolved. We addressed these questions by analyzing the evolution of the key players constituting the cytokinin signaling pathway using the genomes and/or EST collections of key species of bacteria, unicellular eukaryotes, algae, and land plants. Our analysis revealed a previously unknown subfamily of cytokinin receptors found only in early diverging land plants. Although their domain architecture is similar to the architecture of those receptors described in higher land plants, the sequence similarity of residues critical for structure or cytokinin binding of the CHASE domain is comparatively low. Nevertheless, various functional experiments showed the ability of two members of this subfamily to bind cytokinin and translate the binding of different types of cytokinins into a cellular signal. Furthermore, we found hints for the presence of cytokinin receptors in the charophyte algal species *Spirogyra pratensis*, a member of a group of green algae giving rise to land plants. In addition, the analysis revealed the presence of RRBs, which display diverse domain architectures. These and other findings indicate a much greater level of complexity for the evolution of cytokinin signaling than previously anticipated (Pils and Heyl, 2009).

RESULTS

A Subfamily of Cytokinin Receptors Emerged

In the first step of our analysis, we focused on the evolution of the CHASE domains. Both Maximum Likelihood (ML) and Bayesian interference clearly distinguished three different subclades (Fig. 1). Although most of the land plant CHASE domains clustered similarly to what was published previously (Pils and Heyl, 2009), a unique clade containing eight sequences from *Physcomitrella patens* and one sequence from *Marchantia polymorpha* previously not associated with cytokinin signaling emerged. In addition, a third clade contained CHASE domains from receptors of cyanobacteria, *Dictyostelium discoideum* and chlorophyte algae (Fig. 1A). We analyzed all protein sequences in more detail. The observed

domain architecture was different between the three clades, with the two land plant branches showing the conserved domain pattern of cytokinin receptors (CHASE, Histidine Kinase A domain [HisK]; Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase domain [HATPase], and RR domains; Fig. 1B). In contrast, the proteins of the third branch only had the CHASE domain in common but otherwise, displayed a rather diverse array of different domains, such as Guanylate Cyclase or Phosphodiesterase. When we looked at the conservation of those residues of the CHASE domain that were shown to be important for its structure and cytokinin binding (Heyl et al., 2007; Hothorn et al., 2011), a high level of conservation was found among the classical cytokinin receptors, whereas members of the newly identified subfamily had only a very low level of conservation compared with the CHASE domain of *Arabidopsis histidine kinase4* (AHK4; Fig. 1B). To investigate the distinct clade of putative cytokinin receptors in more detail, we checked whether these genes are expressed by looking for EST evidence from *P. patens* itself as well as the closely related moss *Funaria hygrometrica*. For three of these genes, EST data were found from both moss species, and for another four genes, expression evidence came from *P. patens* exclusively (Fig. 1A). This result indicates that most members of this unique family are, indeed, expressed.

To test experimentally whether members of this clade can function as cytokinin receptors, we selected two proteins to serve as examples. *M. polymorpha* CHASE domain containing His kinase receptor1 (MpCHK1) is the only detected putative cytokinin receptor from *M. polymorpha*, the earliest diverging land plant in this study. For the three PpCHKs for which we found EST evidence for their expression in both analyzed moss species, one (PpCHK4) was randomly chosen for additional analysis. Both proteins, PpCHK4 and MpCHK1, were expressed in *Escherichia coli* and tested in a cytokinin binding assay (Mizuno and Yamashino, 2010). As a positive control, we used the cytokinin receptor AHK4, and as a negative control, we used the AHK4 (T301I) mutation, which was originally identified as *wooden leg* (*wol*; Mähönen et al., 2000) and shown to be unable to bind cytokinin (Yamada et al., 2001; Heyl et al., 2007). The assay showed binding of the moss receptor PpCHK4 for trans-zeatin. For MpCHK1, the cytokinin binding was weaker than in the case of PpCHK4 but clearly stronger than the binding detected for the negative control, AHK4 (T301I; Fig. 2). However, also because of different protein expression levels, comparisons of binding levels between the different receptors are difficult (Fig. 2B).

However, to function as a receptor, a protein must be able to translate the binding of the ligand into a cellular signal. Therefore, we used a bacterial complementation system to verify the functionality of the newly identified receptors (Mizuno and Yamashino, 2010). The two potential cytokinin receptors activated the reporter gene specifically in response to various cytokinins in a dose-dependent manner. Trans-zeatin treatment resulted in the strongest activation of the reporter gene for all three functional receptors. Interestingly, although cis-zeatin

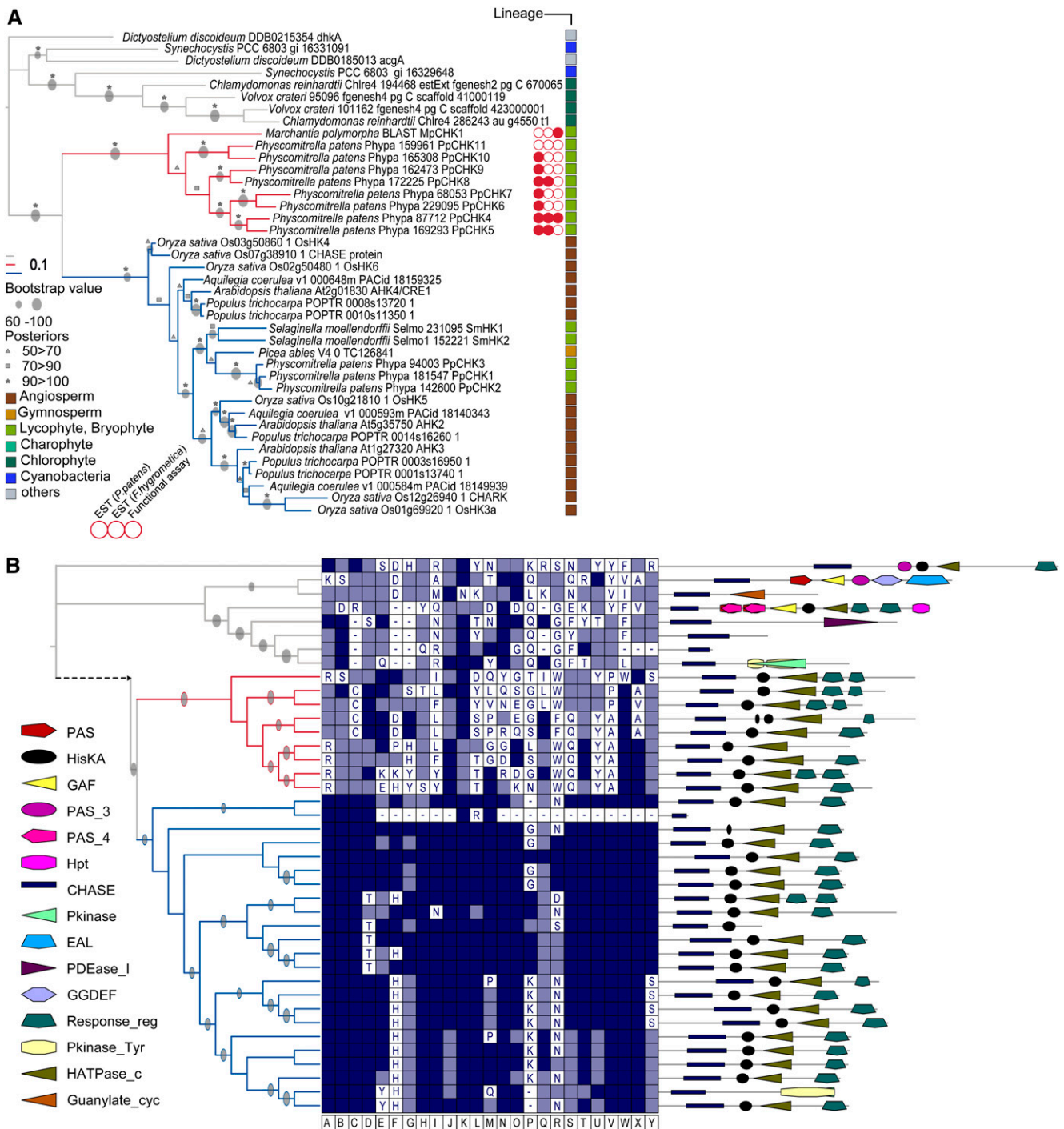


Figure 1. Phylogenetic tree of CHASE domain containing His kinases reveals a novel subfamily of plant cytokinin receptors. A, The CHASE domain sequences from all organisms under investigation were aligned, and a phylogenetic tree using ML and Bayesian methods was inferred (details in “Materials and Methods”). Because tree topology was identical with both methods, bootstrap values and posterior probabilities are depicted on the ML tree. The newly identified subfamily of plant cytokinin receptors is indicated with red lines, and the clade with classical cytokinin receptors is marked by blue lines. EST evidence is symbolized by a red circle, and the lack of EST evidence is symbolized by a white circle. B, Center, Conservation of amino acids important for the structure and function of the CHASE domain (Hothorn et al., 2011). The highlighted positions (A–Y) are in respect to the CHASE domain of AHK4 (details in Supplemental Table S1). Dark blue shows the identity to AHK4, light blue marks conservative substitution, and white boxes symbolize no conservation related to the respective position in the CHASE sequence of AHK4. B, Right, Domain architecture of the whole protein. Details are in “Materials and Methods.” Abbreviations of the domains are according to the Pfam-database (<http://pfam.sanger.ac.uk>): HisKA, PF00512; GGDEF, PF00990;

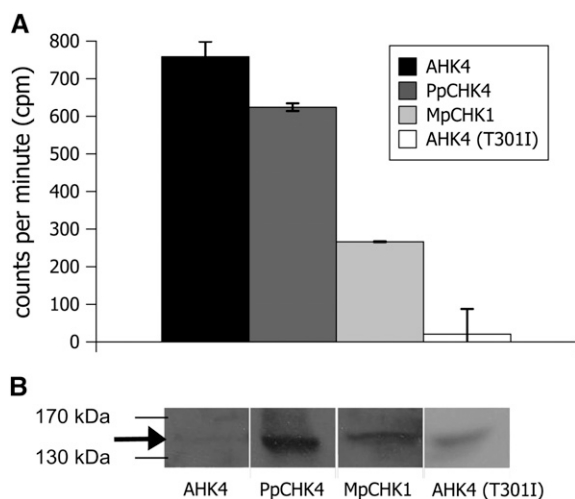


Figure 2. Two members of the newly identified subfamily of putative cytokinin receptors bind cytokinin. A, In vitro binding of trans[2-³H]zeatin to AHK4, PpCHK4, MpCHK1, and AHK4 (T301I) proteins overexpressed in the *E. coli* strain KMI002. Bacterial cells were assayed for specific trans[2-³H]zeatin binding. B, Protein blot of the respective proteins expressed in *E. coli* as detected by glutathione S-transferase antibody. The arrow highlights the band for AHK4.

caused the weakest response of all tested cytokinins in AHK4 of Arabidopsis, it triggered a stronger activation than, for example, benzyladenine (BA) in the moss receptor PpCHK4. This finding might indicate different ligand binding properties for the different receptors. Adenine, which is structurally similar to cytokinin but biologically not active, did not trigger a response by any of the receptors tested. These data show that at least two members of this distinct clade of putative cytokinin receptors have the biochemical characteristics consistent with their domain architecture and thus, are fulfilling the prerequisite for functioning as cytokinin receptors (Fig. 3).

EST Data Point at Algal Origin of the Cytokinin Receptors

In the case of the cytokinin receptors, the inclusion of EST data in the phylogenetic analysis revealed the first evidence, to our knowledge, for an algal origin of cytokinin receptors. We generated phylogenetic trees for each domain of the cytokinin receptor individually to counteract potential bias from using EST data. Regardless of whether the RR (Supplemental Fig. S1), HATPase, or HisK domain was used to generate a phylogenetic tree, we observed similar relationships in the data set (Supplemental Fig. S2). Sequences that clustered together for one domain were also found in the same subclades for the other domains.

Clearly recognizable subfamilies of His kinases were those subfamilies grouping with the osmosensor AHK1 (Arabidopsis; Tran et al., 2007), the sensor kinases Cytokinin independent1 (Arabidopsis; Kakimoto, 1996) and AHK5 (Arabidopsis; Iwama et al., 2007), the different phytochrome kinases (Mathews, 2005), the ethylene receptor clade (Bleecker et al., 1998), and the cytokinin receptors (Heyl et al., 2012). Within the RR tree, the clade representing the His kinases with a CHASE domain contained four EST sequences that have only an RR domain (Supplemental Fig. S1). These sequences originated from the gymnosperm *Picea abies* and the charophyte alga *S. pratensis* and might be parts of complete cytokinin receptors. This hypothesis is further supported by the clustering of an HATPase domain from an *S. pratensis* EST in the clade of the HATPase domain of cytokinin receptors (Supplemental Fig. S2). Thus, the incorporation of EST data in the analysis provided hints at an origin of the cytokinin receptors in charophycean algae outside of the land plants.

Evolution of the HPT Proteins Recapitulates Plant Evolution

The analysis of the HPT proteins resulted in a phylogenetic tree resembling the recently described course of plant evolution (Wodniok et al., 2011; Fig. 4). This result was found regardless of the method used for tree calculation (ML or Bayesian interference). The analysis of the HPTs showed that members of this protein family from bacteria were distinct from those members of algae or land plants not only by their position within the tree but also by the additional domains found in those proteins. In contrast, with one single exception, all HPTs from the green lineage did not contain any additional domains, such as cGMP-specific phosphodiesterases, adenylyl cyclases, and FhlA or CheY-like response regulator domains (Fig. 4). For Arabidopsis, it was shown that one of the HPTs (AHP6) does not contain the canonical His residue and thus, cannot be phosphorylated. However, it functions as a negative regulator of cytokinin signaling (Mähönen et al., 2006a). An analysis for the presence of the conserved His residue in the HPT domains revealed that only angiosperms and gymnosperms contained such noncanonical HPTs, because no proteins carrying such a mutation were found in any of the other clades (Fig. 4 and Supplemental Fig. S3).

RRs Show Distinct Origins

The phylogenetic analysis of the RR domain revealed the RRAs and the RRBs to be closely related to each other,

Figure 1. (Continued.)

Response_reg, PF00027; PAS, PF00989; Pkinase, PF00069; PAS_3, PF08447; Hpt, PF01627; EAL, PF00563; HATPase_c, PF02518; GAF, PF01590; CHASE, PF03924; Pkinase_Tyr, PF07714; PAS_4, PF08448; Guanylate_cyc, PF00211; and PDEase_I, PF00233.

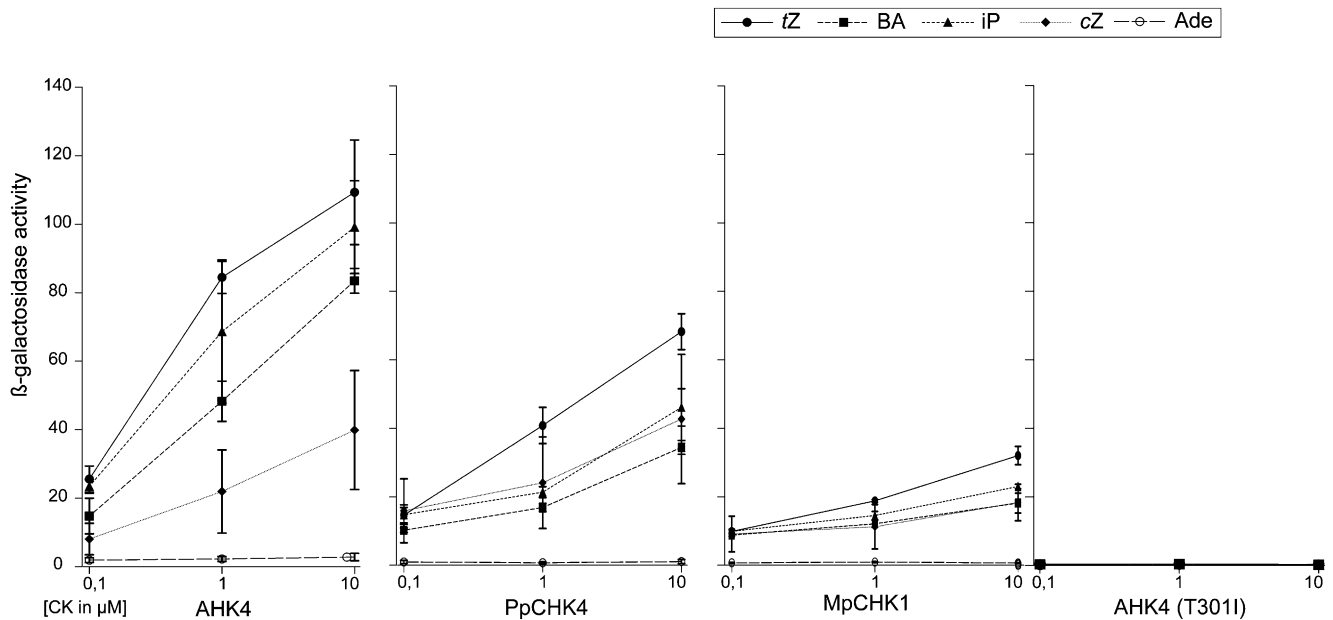


Figure 3. Experimental evidence for a function as cytokinin receptors for two members of the newly identified subfamily. AHK4, PpCHK4, MpCHK1, and AHK4 (T301I) were tested in a bacterial complementation assay with different cytokinin and adenine concentrations. Ade, Adenine; BA, benzyladenine; cZ, cis-zeatin; iP, isopentenyl; tZ, trans-zeatin.

whereas the RRCs grouped to the clade of the RR domains of His kinases (Supplemental Fig. S1). RRCs consist only of the RR domain and thereby, resemble the RRAs in the domain architecture (Mizuno, 2004). However, in contrast to the RRAs, RRCs are not inducible by cytokinin (Kiba et al., 2004). Expression analysis for the two RRCs of *Arabidopsis* revealed a specific expression in different organs of the inflorescence, indicating a role in development (Gattolin et al., 2006). Although mutant analysis revealed no function for this protein family in cytokinin signaling, the detection of interactions of RRCs with different HPT proteins might hint at a role in a TCS (Horák et al., 2008). All RR domains from the data set were analyzed in one phylogenetic tree (Supplemental Fig. S1). This tree showed that the RR domains of the RRCs are more closely related to RR domains of His kinases from plants than the RRAs or RRBs. This finding is in accordance with previous analyses (Kiba et al., 2004; Schaller et al., 2008). Another hint that this class of RRs might be, in fact, degenerated receptors rather than bona fide RRs comes from the presence of an HATPase domain in two members of this clade (Fig. 5). Although the molecular mechanism of RRCs in planta is not known, it is noteworthy that many members of this family showed changes at the canonical Asp position in the DDK motif (Fig. 5).

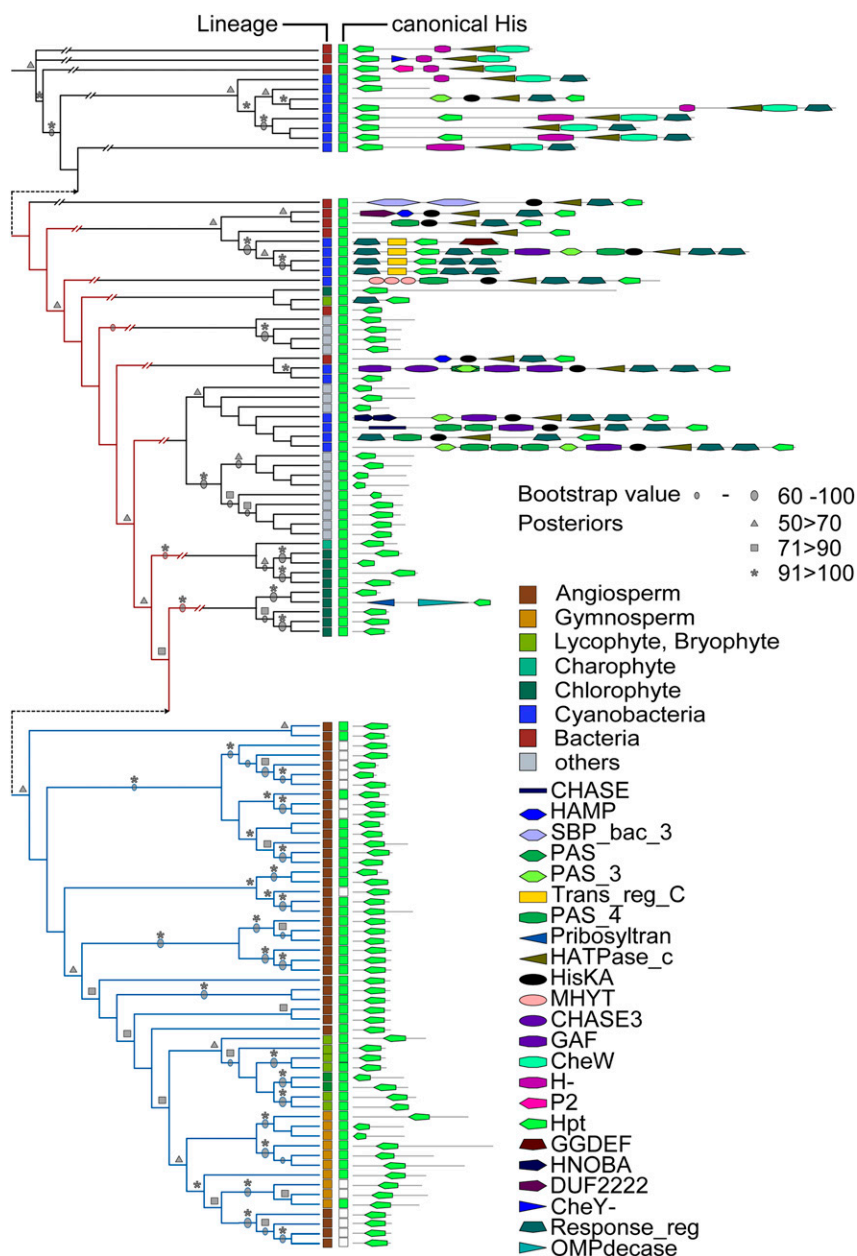
In contrast to the RRCs, the RR domains of the RRAs and RRBs of plants formed a distinct clade. This clade was further subdivided into three subclades (Fig. 6). Although the individual branches differed depending on the phylogenetic method (ML or Bayesian interference), the three major subclades (RRA, RRB, and PRR) were supported by both methods (Supplemental Fig. S4). All known RRAs were found in one monophyletic group with sequences

exclusively from land plants. In almost all of the sequences, the canonical Asp was conserved; however, in three cases (intriguingly restricted to spruce [*P. abies*]), this critical residue was substituted. The other subclade was further subdivided into the RRB branch and a branch containing the so-called PRRs (Makino et al., 2000). PRRs often contain additional domains apart from the RR and the Myb domains, such as a CONSTANCE, CO-like, and True Oscillator Component1 (CCT) domain, and function in circadian rhythm (Mizuno, 2004). Our phylogenetic analysis shows a clear separation between the RRBs and the PRRs, which was also reflected in the domain structure, because most PRRs also contain a CCT or a WD40 repeat domain, neither of which is found among the RRBs. Interestingly, there were a number of sequences in the clade of the RRBs where the DNA binding Myb domain and/or the canonical Asp of the RR domain was missing (Fig. 6). Thus, they are lacking the key features of this class of transcription factors for the mediation of the transcriptional response to cytokinin and therefore, might not be functional in that capacity. Similar results have recently been reported for rice (Tsai et al., 2012). In the clade of the PRRs, a curiously high number of proteins, especially those proteins from algae and basal land plants, shared the conservation of the canonical Asp of the RR domain with the bona fide RRs and thus, could be functional in a TCS context. In contrast, all the angiosperm sequences were missing the canonical Asp residue (Fig. 6).

DISCUSSION

Understanding the principles and mechanisms leading to evolutionary innovations is a central theme in biology.

Figure 4. Phylogenetic tree for the different HPT proteins investigated in this study. Left, Sequences of the HPT domain were aligned, and a phylogenetic tree using ML or Bayesian methods was inferred (black lines). Both trees show identical topology; therefore, bootstrap and posterior probabilities are depicted on the tree. Protein identifiers are shown in Supplemental Figure S3. The clade containing proteins involved in cytokinin signaling is highlighted with blue lines. For canonical His, green indicates conserved His residue on relevant position, and white boxes do not show any conservation. Right, Domain architecture of the whole respective protein. Details are in "Materials and Methods." Abbreviations of the domains are according to the Pfam-database (<http://pfam.sanger.ac.uk>): HisKA, PF00512; CHASE, PF03924; SBP_bac_3, PF00497; GAF, PF01590; Response_reg, PF00027; P2, PF07194; Trans_reg_C, PF00486; CheY-, CL0304; CHASE3, PF05227; HAMP, PF00672; PAS_3, PF08447; CheW, PF01584; GGDEF, PF00990; HNOBA, PF07701; Pribosyltran, PF000156; OMPdecase, PF00215; MHYT, PF03707; PAS, PF00989; PAS_4, PF08448; H-, PF02895; Hpt, PF01627; DUF2222, PF09984; and HATPase_c, PF02518.



With the increasing amount of sequence information available, phylogenomic analysis and the investigation of the evolution of protein families or whole pathways that are necessary for the investigation of evolutionary novelties have become possible. Taking advantage of these new resources, the aim of this study was to use a comprehensive phylogenomic approach to investigate how cytokinins, ubiquitous degradation products of nucleic acids, evolved into critical signaling molecules for plants, going far beyond the scope of our previous analysis (Pils and Heyl, 2009).

Results of Phylogenetic Analysis Mirror Known Functional Relations

The necessity of using a large sample set to include the whole entity of proteins harboring the domain of interest

weakens the detectable phylogenetic signal. This effect is amplified by the restriction of our analysis to the conserved regions of the proteins. Nevertheless, for example, in the case of the receptors, we analyzed the receptors domains independently and found very similar functional clustering in all four independent trees (Supplemental Fig. S2). For all phylogenetic trees that were constructed, we also analyzed the domain composition of the whole protein, showing that the published biological function of the proteins is well-reflected in the phylogenetic signal obtained by their single domains. Furthermore, available biological evidence on the function of different proteins in the respective clades supports our phylogenetic results. We, therefore, conclude that the reliability of the depicted phylogenetic trees is not solely because of their probability of

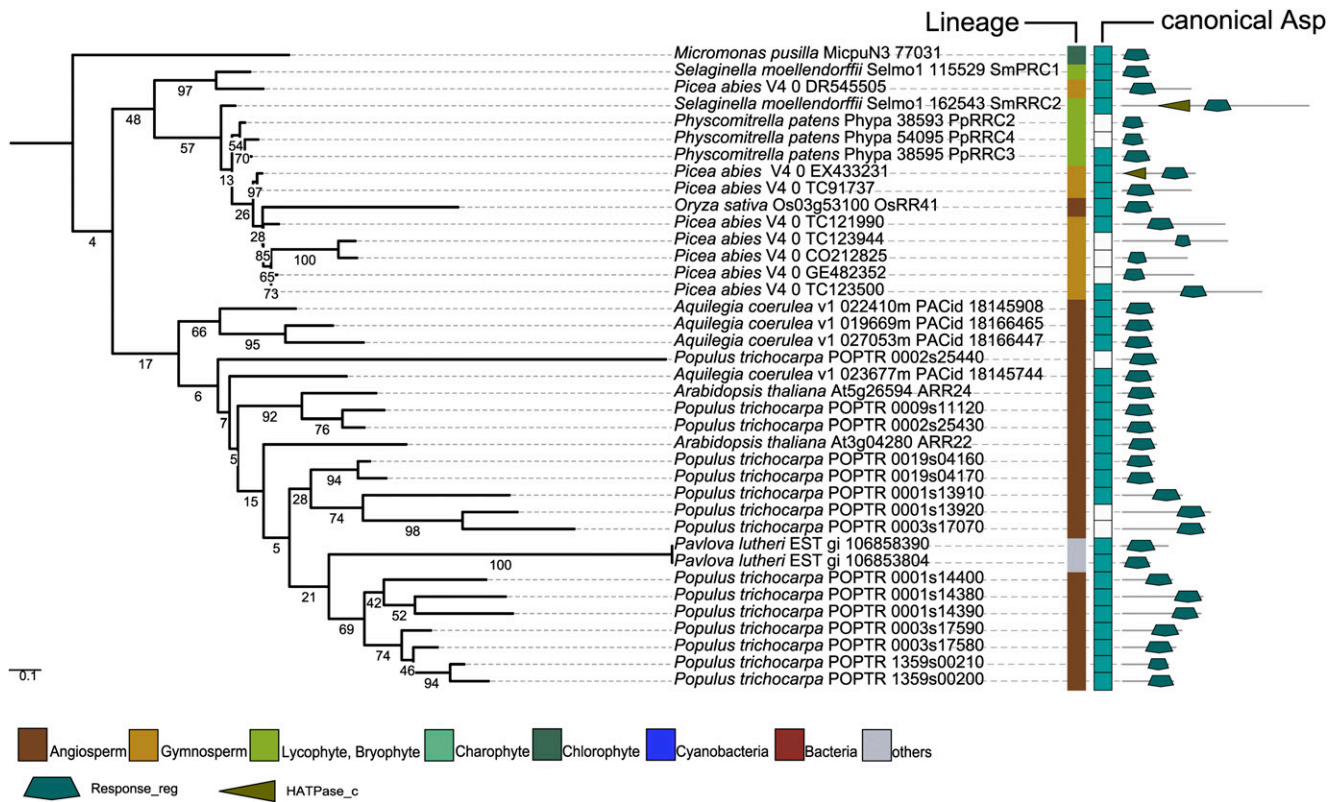


Figure 5. Phylogenetic tree for the RRC. Left, The sequences of the phylogenetic subtree of the RRC clade identified previously (Supplemental Fig. S1) were aligned, and an ML tree was calculated. For canonical Asp, green indicates conserved Asp residue on relevant position, and white boxes do not show any conservation. Right, Domain architecture of the whole respective protein. Details are in "Materials and Methods." Abbreviations of the domains are according to the Pfam-database (<http://pfam.sanger.ac.uk>): Response_reg, PF00027; HATPase_c, PF02518.

occurrence but also, in agreement with the available biological data.

Cytokinin Perception through the CHASE Domain Might Have Emerged Shortly before the Conquest of Land

One of the most surprising results of this study was the detection and validation of a previously unknown subfamily of cytokinin receptors, which were shown to bind the phytohormone and translate this binding into a cellular response. At this point, we can only speculate about a possible biological function of members of this unique subfamily. The fact that only members of *P. patens* and *M. polymorpha* are present in this subclade hints to an early separation of this group of cytokinin receptors from those receptors also found in modern land plants. There are three additional cytokinin receptors from *P. patens* clustering with the classical cytokinin receptors from the other land plants. In contrast, MpCHK1 from *M. polymorpha* is the only cytokinin receptor found in this species. However, it could be because either we had limited access to the genome of this liverwort (we obtained only BLAST results from the genomic sequence using *AHK4* of *Arabidopsis* as a query) or there is only one cytokinin receptor

in this species. Thus, there might be more cytokinin receptors in the *M. polymorpha* genome that have not yet been identified. Only the sequencing of more genomes from charophyte algae or species at the base of land plant evolution will allow us to identify the last common ancestor of these two receptor subfamilies. For such an analysis, the region of the CHASE domain might be enlarged, because recent analysis found high sequence conservation in the amino acids on both sides of the CHASE domain (Steklov et al., 2013).

Another surprising result of this analysis was the detection of two ESTs in the RR tree and one EST in the HATPase tree of the charophyte alga *S. pratensis* that clustered with those sequences of the cytokinin receptors. These results suggest that those ESTs might be part of cytokinin receptors or ancestral receptors. It will be interesting to investigate whether the respective proteins function in cytokinin perception. The analysis also confirmed that the RRCs clustered with the hybrid His kinases rather than with the other RRs (Kiba et al., 2004; Schaller et al., 2008). Remarkably, some RRCs and also, some hybrid His kinases share a phosphatase activity and thus, might work as negative regulators of TCS pathways (Kiba et al., 2004; Mähönen et al., 2006b; Horák et al., 2008). Additional evidence for

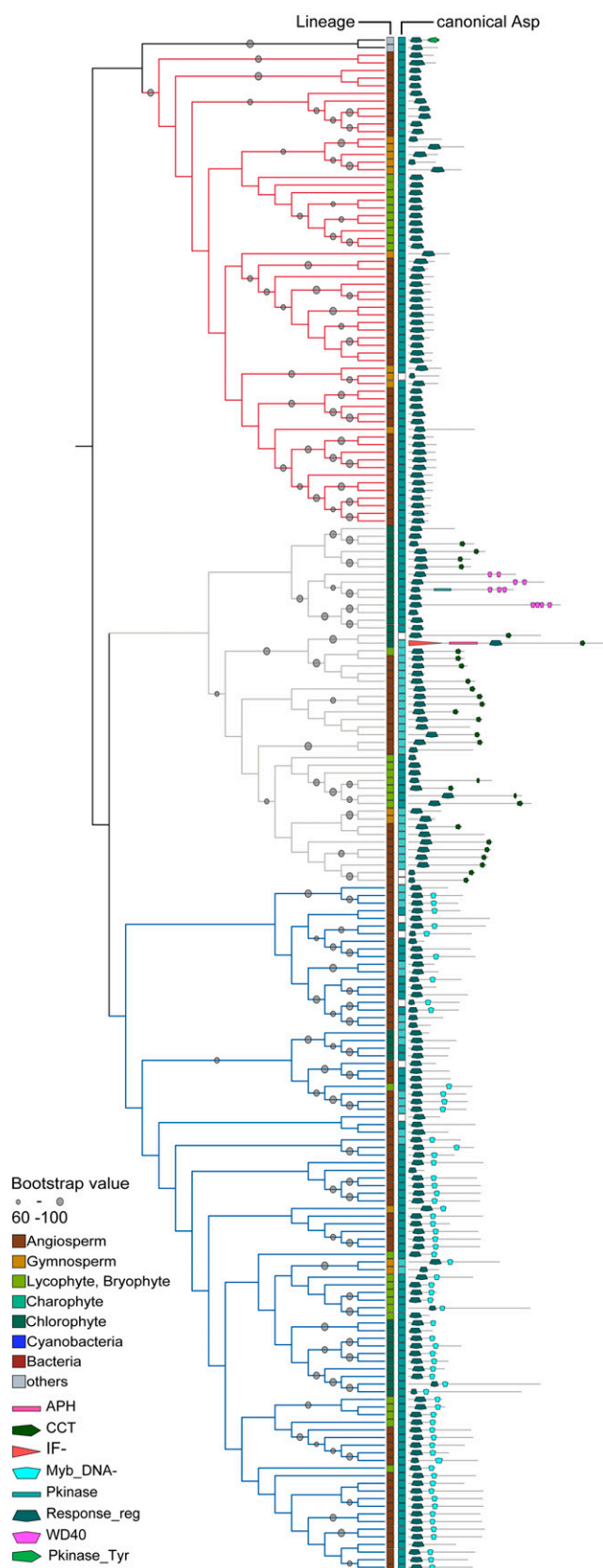


Figure 6. Phylogenetic relationship of RRAs, RRBs, and PRRs. Left, Sequences of the RR domains were aligned, and an ML tree was

RRCs as degenerated hybrid His kinases comes from the finding that two members of this clade also contain HATPase domains, which are part of this receptor type. Thus, it might be sensible to include the RRCs in the context of cytokinin receptor evolution. Thus, it is worth mentioning that the overexpression of one of the RRCs from *Arabidopsis*, ARR22, leads to a similar phenotype as the *wol* mutation of the cytokinin receptor mutant AHK4 (Kiba et al., 2004). Again, more sequence data from charophycean algae and early divergent land plants as well as more information about their biological function will be necessary to shed light on the evolution of this protein family.

Phylogenetic Analyses of HPT Proteins Point at Changing Functions during the Course of Evolution

The phylogenetic tree of the HPT proteins indicates a tendency in which the bacterial HPT domains are part of large multidomain proteins (e.g. receptors), whereas in unicellular eukaryotes and plants, the HPT domain is the only domain of a short protein, which could point to different molecular functions of the HPT domain containing proteins. In eukaryotes, the nuclear membrane separates the cytoplasmic part of receptors from nuclear RRs. Thus, the need arises for HPTs that can easily shuttle between the two compartments, which was shown for *Arabidopsis* (Punwani et al., 2010). In contrast, in bacteria, which lack the nuclear membrane, RRs can interact directly with the His kinase receptors. Thus, there is no need for membrane-crossing HPTs.

Another important result from the phylogenetic analysis of the HPT sequences concerns the canonical His residue. Although this critical amino acid is a prerequisite for a functioning in the His-to-Asp phosphorelay, it has also been shown that family members lacking it can play a role in cytokinin signaling (Mähönen et al., 2006a). In our analysis, noncanonical HPTs were restricted to vascular plants, where they might add an additional layer of regulation necessary for correct execution of the more complex developmental programs.

RRAs and RRBs Share a Common Origin

The results of the phylogenetic analysis revealed a different picture for the nonreceptor types of RRs. The RRBs are plant specific and were found throughout the plant kingdom. This finding, together with the fact that many

calculated. Depicted is a subtree (protein identifiers are shown in Supplemental Fig. S5) from the global RR tree (Supplemental Fig. S1) containing the RRAs (red lines), RRBs (blue lines), and PRRs (gray lines). For canonical Asp, dark blue shows conservation of canonical Asp, light blue marks conservative substitution, and white boxes symbolize no conservation. Right, Domain architecture of the whole respective protein. Details are in "Materials and Methods." Abbreviation of the domains are according to the Pfam-database (<http://pfam.sanger.ac.uk>): Pkinase, PF00069; APH, PF01636; Response_reg, PF00027; CCT, PF06203; Myb_DNA-, PF00249; WD40, PF00400; and IF-, PF01008.

RRBs were missing either the Myb-related DNA binding domain and/or the canonical Asp residue of the RR domain, might indicate that they represent an evolutionarily older protein family compared with the RRAs or that the selective force acting on them is lower than on RRAs. However, even the noncanonical members of this clade are phylogenetically clearly distinct from the PRRs (Makino et al., 2000). The branching pattern of the RR tree hints at an early split in the evolution of these two subclades. Interestingly, it was shown that some PRRs have similar function as regulators of the circadian rhythm in *Chlamydomonas* spp. and the land plant *Arabidopsis* (Mizuno, 2004; Matsuo et al., 2008). In contrast, nothing is known about the function of the RRBs in algae, whose genomes do not encode cytokinin receptors. In contrast, sequences for RRAs have been found neither in charophyte algae nor among the chlorophyte algae. One explanation could be that the last common ancestor of land plants and algae still had RRAs, but they were eventually lost in the algal species, whereas in land plants, these proteins acquired a new function as regulators in the newly established cytokinin signaling pathway and were subsequently conserved. Alternatively, their separation from the last common ancestor of RRAs and RRBs happened after the split of the Chlorophyceae and the Charophyceae.

CONCLUSION

In the era of high-throughput sequencing and the resulting availability of a plethora of genomic sequences, phylogenetic analysis has become a basic component in the characterization of a given gene. However, such results can be misleading. This study highlights the importance of looking at not just the evolutionary history of a single gene but the whole associated pathway. In addition, the results of our analysis, especially the identification of a unique subfamily of cytokinin receptors, vindicate the use of a domain-based approach to identify members of protein families, especially when the data include ESTs.

Taking these guiding principles into account, we conducted a comprehensive analysis of the components of the cytokinin signaling pathway. Our analysis showed that all the protein domains necessary for this signaling system are present in bacteria. During the course of plant evolution, these domains were then assembled in such a way that they could be used for signal transduction of the phytohormone cytokinin. Such a model is also supported by a recent analysis of cytokinin metabolism genes (Frébert et al., 2011; Spichal, 2012). This set of evolutionary changes led to the cytokinin regulatory system well characterized in the angiosperms (Gruhn and Heyl, 2013).

MATERIALS AND METHODS

Range of Species Included in the Analysis

Because the aim of this study was to investigate the origin and evolution of the cytokinin regulatory system, a balanced sampling from all relevant phyla was necessary. In total, we included in our analysis the genome and/or EST data of 11

bacterial, 13 unicellular eukaryote, 12 chlorophyte algae, three charophyte algae, and eight land plant species (Supplemental Table S2). Available nucleotide sequences (ESTs) were translated into all six frames using the software virtual ribosome (version 1.1; Wernersson, 2006). For the nomenclature of TCS components, we followed the recently published guidelines (Heyl et al., 2013).

Domain Composition Analysis

Domain profiles of characteristic proteins involved in the cytokinin network were obtained from the public database Pfam (Finn et al., 2010; Supplemental Table S3). The respective proteomes were analyzed for the existence of the relevant domain by Hidden Markov Model (HMM) search of the HMMER3 Package (HMMER3.0; Finn et al., 2011) using the PFAM-A model available on the PFAM Web site (Supplemental Table S2) applying the gathering cutoff option. Subsequently, an HMM search was undertaken on the protein sequences that were identified previously to contain the respective domain using HMMER2 (HMMER2.4i). The HMMER2 domain models were based on the seed alignments initially used to generate the HMMER3 models. The resulting models were subsequently calibrated using hmmlibrate (HMMER2 package). Domains were extracted according to the HMM2 result from the protein sequence and aligned. For multiple sequence alignments (MSAs), the MAFFT package (MAFFT v6.846b; Katoh et al., 2009) using the local pairwise alignment option was applied. For tree construction and bootstrapping, RAxML package (RAxML v7.0.4; Stamatakis, 2006) was used running under the fast optimization method and searched for the best scoring ML tree using the amino acid substitution matrix Whelan and Goldman with gamma model for rate of heterogeneity. Each tree was calculated with 100 bootstraps, and automatic eliminations of redundant sequences were accepted. Alternatively, we inferred phylogenetic trees using MrBayes (<http://mrbayes.sourceforge.net>) with the Whelan and Goldman amino acid substitution matrix, and the γ -shaped rate variation with a proportion of invariable sites was estimated using eight rate categories. Markov Chain Monte Carlo was run with five independent runs of four chains ($T = 0.05$) and a sample frequency of 10 for 5,000,000 (RR and CHASE domain) or 10,000,000 (HPT) generations. At this point, the average SD of split frequency, a measure of convergence, was well below 0.02 for each analysis, and the runs were stopped. The first 25% of samples of each run were discarded as burn-in, and the remaining samples were used to infer tree and branch topology.

Conservation Comparison of Critical Residues from Domains under Investigation

Amino acid residues, which are critical for the function of the investigated proteins, were compared. The characteristic amino acid of the protein in the tree founding alignment was compared with published, characterized residues in known proteins. Because *Arabidopsis* (*Arabidopsis thaliana*) is the best studied organism in relation to cytokinin signaling, *Arabidopsis* proteins were used as template proteins (CHASE domain [Heyl et al., 2007; Hothorn et al., 2011], HisK domain [Ueguchi et al., 2001], HPT domain [Heyl and Schmülling, 2003], and DDK motif of the RR domain [Ueguchi et al., 2001]).

Cytokinin Binding Assay

The cytokinin binding assay was performed as described before (Mizuno and Yamashino, 2010). The cytokinin receptor *Arabidopsis* His kinase AHK4, AHK4 (T301I), the putative cytokinin receptor *Physcomitrella patens* receptor PpCHK4, and the *Marchantia polymorpha* receptor MpCHK1 were cloned into the pDEST15 vector (Invitrogen) and expressed using the *Escherichia coli* strain KMI002 (Mizuno and Yamashino, 2010). AHK4 (T301I) was used as a negative control. Tritium-labeled trans-[³H]zeatin (592 GBq/mmol) was obtained from the Isotope Laboratory of the Institute of Experimental Botany. The coding sequences for MpCHK1 and PpCHK4 were obtained by gene synthesis (Genscript). The protein expression was confirmed by western blot as described before (Heyl et al., 2007).

Semiquantitative Complementation Assay

The functionality of the newly identified cytokinin receptors was tested by a semiquantitative complementation assay as described before (Mizuno and Yamashino, 2010). The *E. coli* strain KMI002 was transformed with the vector pDEST15 (Invitrogen) expressing the *Arabidopsis* His kinase AHK4, AHK4 (T301I), the putative *P. patens* receptor PpCHK4, or the putative *M. polymorpha* receptor MpCHK1.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Phylogenetic tree for the response regulator domain.

Supplemental Figure S2. Phylogenetic subtrees for different domains of hybrid His kinases.

Supplemental Figure S3. Phylogenetic tree for the different His phosphotransfer (HPT) proteins investigated in this study.

Supplemental Figure S4. Phylogenetic relationship of type-A, type-B and pseudo response regulators.

Supplemental Figure S5. Phylogenetic relationship of type-A, type-B and pseudo response regulators (complete identifiers).

Supplemental Table S1. Important residues in the CHASE domain of AHK4.

Supplemental Table S2. Source for the different genomes and ESTs used in this study.

Supplemental Table S3. Pfam profiles and abbreviations for the different domains investigated.

ACKNOWLEDGMENTS

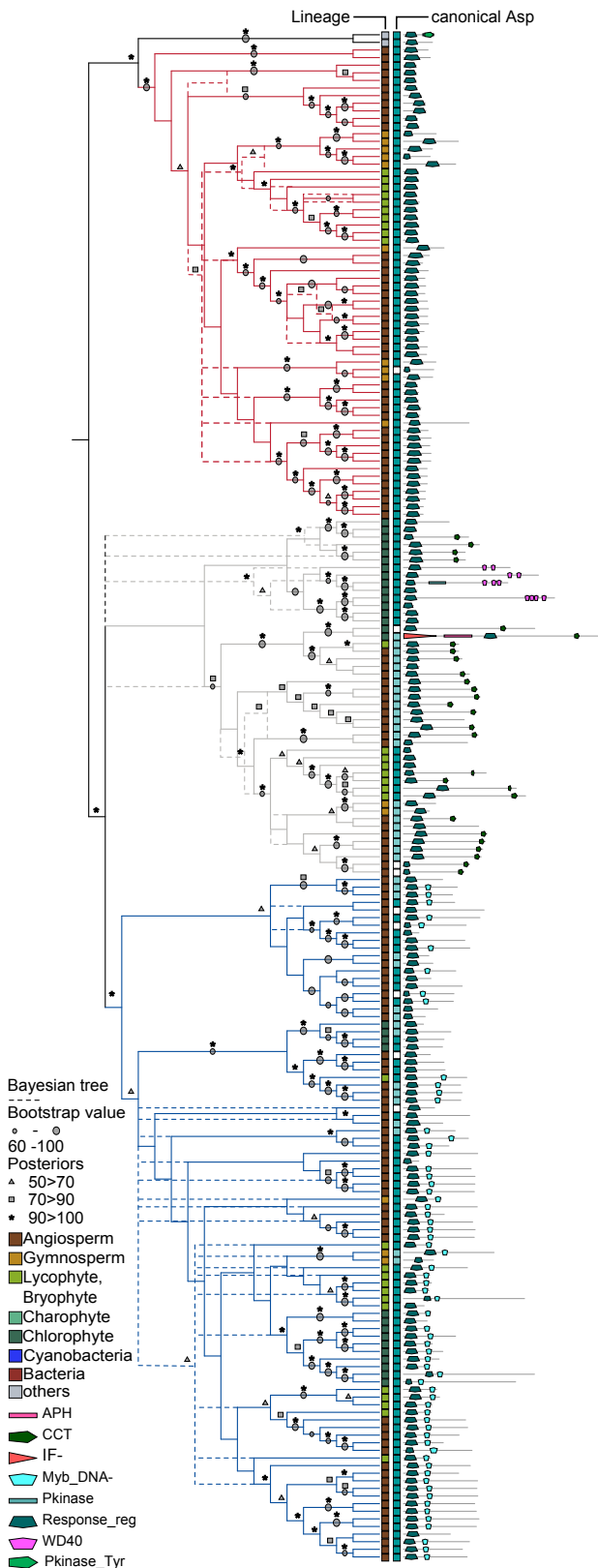
We thank Sandra K. Floyd and John Bowman (Monash University and the U.S. Department of Energy Joint Genome Institute, which is supported by the Office of Science of the U.S. Department of Energy [contract no. DE-AC02-05CH1123]) for sharing the results of BLAST searches of the unpublished *M. polymorpha* genome with us.

Received October 15, 2013; accepted February 7, 2014; published February 11, 2014.

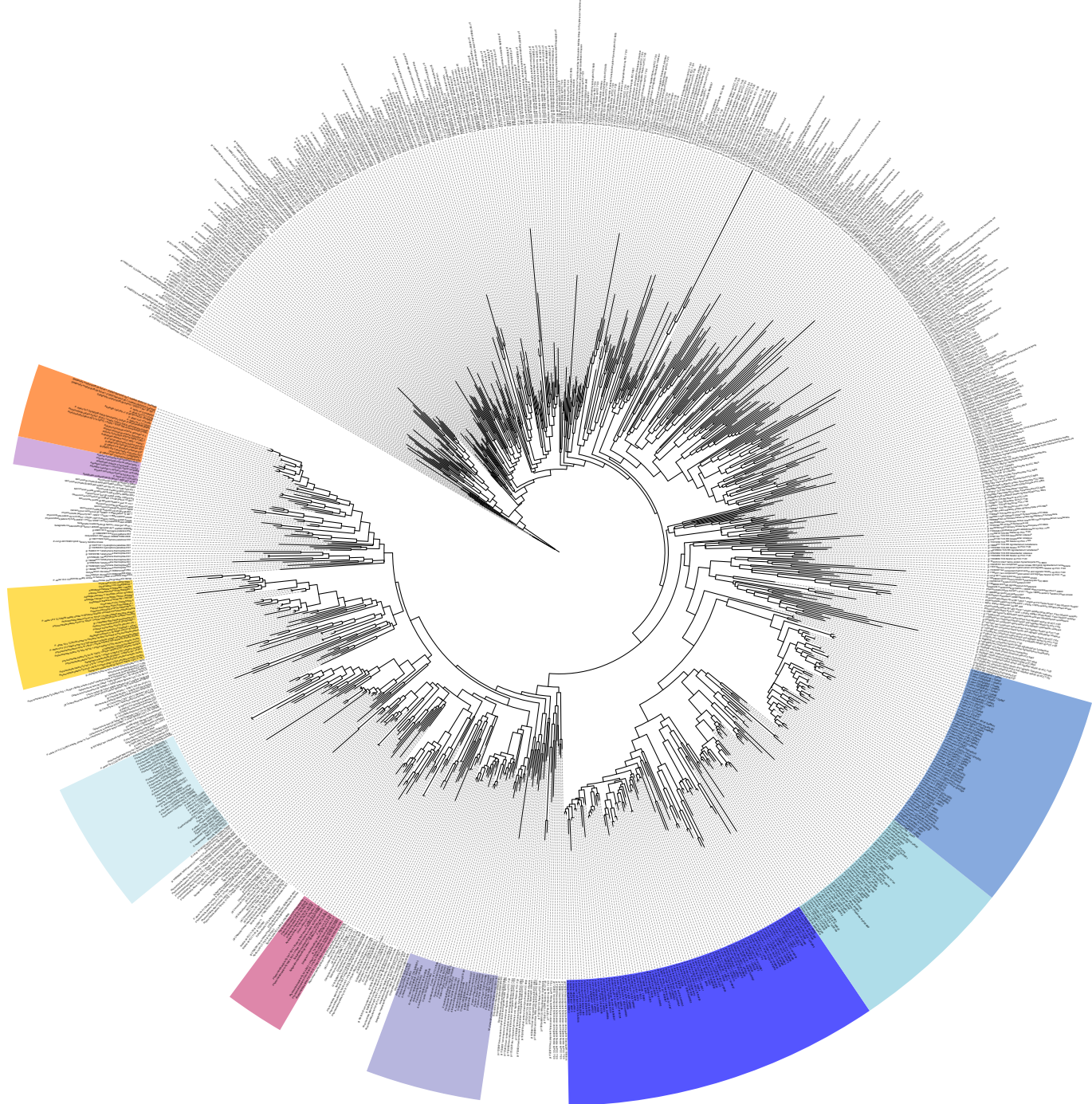
LITERATURE CITED

- Anantharaman V, Aravind L (2001) The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. *Trends Biochem Sci* 26: 579–582
- Argueso CT, Ferreira FJ, Kieber JJ (2009) Environmental perception avenues: the interaction of cytokinin and environmental response pathways. *Plant Cell Environ* 32: 1147–1160
- Bleecker AB, Esch JJ, Hall AE, Rodríguez FI, Binder BM (1998) The ethylene-receptor family from *Arabidopsis*: structure and function. *Philos Trans R Soc Lond B Biol Sci* 353: 1405–1412
- Brenner WG, Ramireddy E, Heyl A, Schmölling T (2012) Gene regulation by cytokinin in *Arabidopsis*. *Front Plant Sci* 3: 8
- Choi J, Choi D, Lee S, Ryu CM, Hwang I (2011) Cytokinins and plant immunity: old foes or new friends? *Trends Plant Sci* 16: 388–394
- Cock JM, Coelho SM (2011) Algal models in plant biology. *J Exp Bot* 62: 2425–2430
- El-Showk S, Ruonala R, Helariutta Y (2013) Crossing paths: cytokinin signalling and crosstalk. *Development* 140: 1373–1383
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–W37
- Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211–D222
- Frébort I, Kowalska M, Hluska T, Frébortová J, Galuszka P (2011) Evolution of cytokinin biosynthesis and degradation. *J Exp Bot* 62: 2431–2452
- Gattolin S, Alandete-Saez M, Elliott K, Gonzalez-Carranza Z, Naomab E, Powell C, Roberts JA (2006) Spatial and temporal expression of the response regulators ARR22 and ARR24 in *Arabidopsis thaliana*. *J Exp Bot* 57: 4225–4233
- Gruhn N, Heyl A (2013) Updates on the model and the evolution of cytokinin signaling. *Curr Opin Plant Biol* 16: 569–574
- Ha S, Vankova R, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS (2012) Cytokinins: metabolism and function in plant adaptation to environmental stresses. *Trends Plant Sci* 17: 172–179
- Hellmann E, Gruhn N, Heyl A (2010) The more, the merrier: cytokinin signaling beyond *Arabidopsis*. *Plant Signal Behav* 5: 1384–1390
- Heyl A, Brault M, Frugier F, Kuderova A, Lindner AC, Motyka V, Rashotte AM, Schwartzberg KV, Vankova R, Schaller GE (2013) Nomenclature for members of the two-component signaling pathway of plants. *Plant Physiol* 161: 1063–1065
- Heyl A, Riefler M, Romanov GA, Schmölling T (2012) Properties, functions and evolution of cytokinin receptors. *Eur J Cell Biol* 91: 246–256
- Heyl A, Schmölling T (2003) Cytokinin signal perception and transduction. *Curr Opin Plant Biol* 6: 480–488
- Heyl A, Wulfetange K, Pils B, Nielsen N, Romanov GA, Schmölling T (2007) Evolutionary proteomics identifies amino acids essential for ligand-binding of the cytokinin receptor CHASE domain. *BMC Evol Biol* 7: 62
- Horák J, Grefen C, Berendzen KW, Hahn A, Stierhof YD, Stadelhofer B, Stahl M, Konz C, Harter K (2008) The *Arabidopsis thaliana* response regulator ARR22 is a putative AHP phospho-histidine phosphatase expressed in the chalazal of developing seeds. *BMC Plant Biol* 8: 77
- Hosoda K, Imamura A, Katoh E, Hata T, Tachiki M, Yamada H, Mizuno T, Yamazaki T (2002) Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the *Arabidopsis* response regulators. *Plant Cell* 14: 2015–2029
- Hothorn M, Dabi T, Chory J (2011) Structural basis for cytokinin recognition by *Arabidopsis thaliana* histidine kinase 4. *Nat Chem Biol* 7: 766–768
- Hwang I, Sheen J (2001) Two-component circuitry in *Arabidopsis* cytokinin signal transduction. *Nature* 413: 383–389
- Hwang I, Sheen J, Müller B (2012) Cytokinin signaling networks. *Annu Rev Plant Biol* 63: 353–380
- Iwama A, Yamashino T, Tanaka Y, Sakakibara H, Kakimoto T, Sato S, Kato T, Tabata S, Nagatani A, Mizuno T (2007) AHK5 histidine kinase regulates root elongation through an ETR1-dependent abscisic acid and ethylene signaling pathway in *Arabidopsis thaliana*. *Plant Cell Physiol* 48: 375–380
- Jaillais Y, Chory J (2010) Unraveling the paradoxes of plant hormone signaling integration. *Nat Struct Mol Biol* 17: 642–645
- Kakimoto T (1996) *CKII*, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* 274: 982–985
- Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* 537: 39–64
- Kiba T, Aoki K, Sakakibara H, Mizuno T (2004) *Arabidopsis* response regulator, *ARR22*, ectopic expression of which results in phenotypes similar to the *wol* cytokinin-receptor mutant. *Plant Cell Physiol* 45: 1063–1077
- Mähönen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN, Helariutta Y (2000) A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes Dev* 14: 2938–2943
- Mähönen AP, Bishopp A, Higuchi M, Nieminen KM, Kinoshita K, Törmäkangas K, Ikeda Y, Oka A, Kakimoto T, Helariutta Y (2006a) Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* 311: 94–98
- Mähönen AP, Higuchi M, Törmäkangas K, Miyawaki K, Pischke MS, Sussman MR, Helariutta Y, Kakimoto T (2006b) Cytokinins regulate a bidirectional phosphorelay network in *Arabidopsis*. *Curr Biol* 16: 1116–1122
- Makino S, Kiba T, Imamura A, Hanaki N, Nakamura A, Suzuki T, Taniguchi M, Ueguchi C, Sugiyama T, Mizuno T (2000) Genes encoding pseudo-response regulators: insight into His-to-Asp phosphorelay and circadian rhythm in *Arabidopsis thaliana*. *Plant Cell Physiol* 41: 791–803
- Mathews S (2005) Phytochrome evolution in green and nongreen plants. *J Hered* 96: 197–204
- Matsuo T, Okamoto K, Onai K, Niwa Y, Shimogawara K, Ishiura M (2008) A systematic forward genetic analysis identified components of the *Chlamydomonas* circadian system. *Genes Dev* 22: 918–930
- Miwa K, Ishikawa K, Terada K, Yamada H, Suzuki T, Yamashino T, Mizuno T (2007) Identification of amino acid substitutions that render the *Arabidopsis* cytokinin receptor histidine kinase AHK4 constitutively active. *Plant Cell Physiol* 48: 1809–1814
- Mizuno T (2004) Plant response regulators implicated in signal transduction and circadian rhythm. *Curr Opin Plant Biol* 7: 499–505
- Mizuno T, Yamashino T (2010) Biochemical characterization of plant hormone cytokinin-receptor histidine kinases using microorganisms. *Methods Enzymol* 471: 335–356

- Mougel C, Zhulin IB (2001) CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends Biochem Sci* **26**: 582–584
- Pils B, Heyl A (2009) Unraveling the evolution of cytokinin signaling. *Plant Physiol* **151**: 782–791
- Punchnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, et al (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carterii*. *Science* **329**: 223–226
- Punwani JA, Hutchison CE, Schaller GE, Kieber JJ (2010) The subcellular distribution of the Arabidopsis histidine phosphotransfer proteins is independent of cytokinin signaling. *Plant J* **62**: 473–482
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, et al (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**: 64–69
- Sakai H, Aoyama T, Oka A (2000) *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. *Plant J* **24**: 703–711
- Schaller GE, Kieber JJ, Shiu SH (2008) Two-component signaling elements and histidyl-aspartyl phosphorelays. *The Arabidopsis Book* **6**: e0112, doi/10.1199/tab.0112
- Spichal L (2012) Cytokinins—recent news and views of evolutionally old molecules. *Funct Plant Biol* **39**: 267–284
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690
- Steklov MY, Lomin SN, Osolodkin DI, Romanov GA (2013) Structural basis for cytokinin receptor signaling: an evolutionary approach. *Plant Cell Rep* **32**: 781–793
- To JP, Haberer G, Ferreira FJ, Deruère J, Mason MG, Schaller GE, Alonso JM, Ecker JR, Kieber JJ (2004) Type-A *Arabidopsis* response regulators are partially redundant negative regulators of cytokinin signaling. *Plant Cell* **16**: 658–671
- Tran LS, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K, Yamaguchi-Shinozaki K (2007) Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis. *Proc Natl Acad Sci USA* **104**: 20623–20628
- Tsai YC, Weir NR, Hill K, Zhang W, Kim HJ, Shiu SH, Schaller GE, Kieber JJ (2012) Characterization of genes involved in cytokinin signaling and metabolism from rice. *Plant Physiol* **158**: 1666–1684
- Ueguchi C, Koizumi H, Suzuki T, Mizuno T (2001) Novel family of sensor histidine kinase genes in *Arabidopsis thaliana*. *Plant Cell Physiol* **42**: 231–235
- Vanstraelen M, Benková E (2012) Hormonal interactions in the regulation of plant development. *Annu Rev Cell Dev Biol* **28**: 463–487
- Werner T, Schmülling T (2009) Cytokinin action in plant development. *Curr Opin Plant Biol* **12**: 527–538
- Wernersson R (2006) Virtual Ribosome—a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res* **34**: W385–W388
- Wodniok S, Brinkmann H, Glöckner G, Heidel AJ, Philippe H, Melkonian M, Becker B (2011) Origin of land plants: do conjugating green algae hold the key? *BMC Evol Biol* **11**: 104
- Yamada H, Suzuki T, Terada K, Takei K, Ishikawa K, Miwa K, Yamashino T, Mizuno T (2001) The *Arabidopsis* AHK4 histidine kinase is a cytokinin-binding receptor that transduces cytokinin signals across the membrane. *Plant Cell Physiol* **42**: 1017–1023



Supplemental Figure S4. Phylogenetic relationship of type-A, type-B and pseudo response regulators. The phylogenetic tree for the sub clades of the RRA, RRB and PRRs generated using a Maximum likelihood algorithm (shown in Fig. 6) is compared to a tree calculated using Bayesian interference. The results for Maximum Likelihood approach is represented by straight lines, the Bayesian interference tree is represented by dashed lines. Bootstrap and posterior values are given.



Color ranges:

Cytokinin

CKI2

CKI1

AHK1

Ethylen

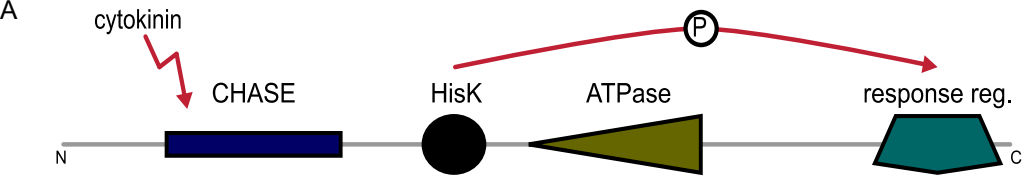
C Type RR

A Type RR

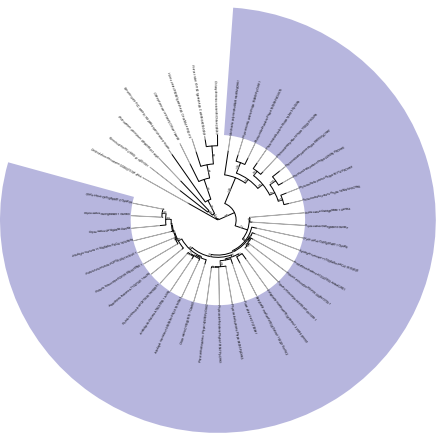
PSEUDO RR

B Type RR

Supplemental Figure S1. Phylogenetic tree for the response regulator domains. All response regulator domains from the dataset were aligned and a Maximum Likelihood tree was calculated.



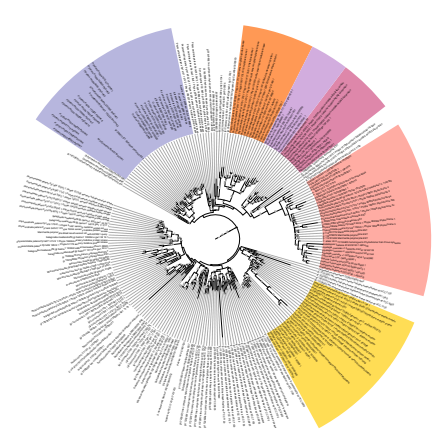
B CHASE domain



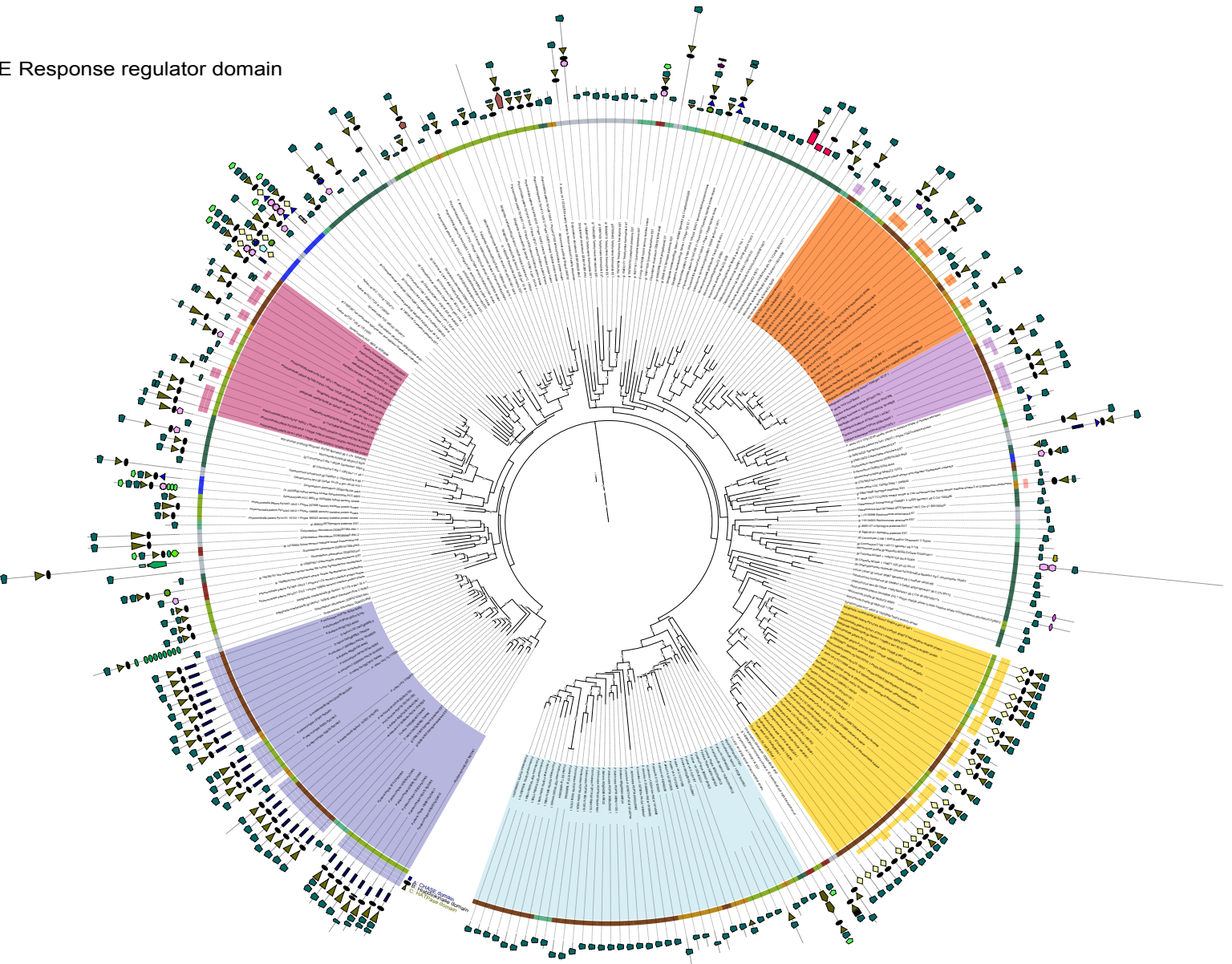
C Histidinkinase domain



D HATPase domain



E Response regulator domain



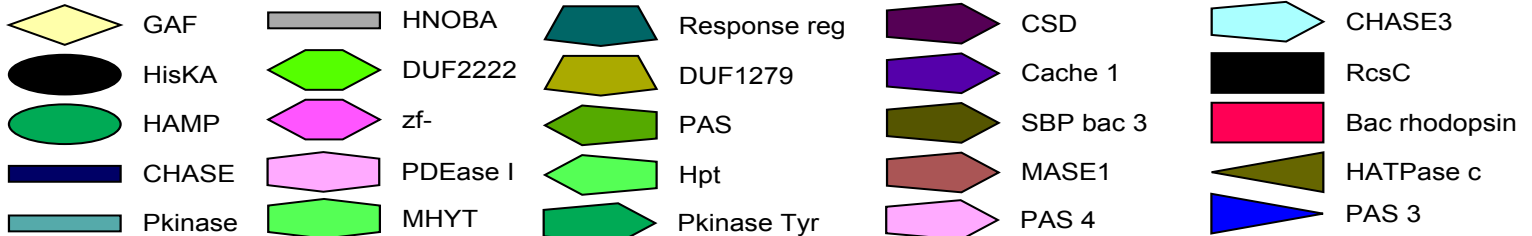
Color ranges:



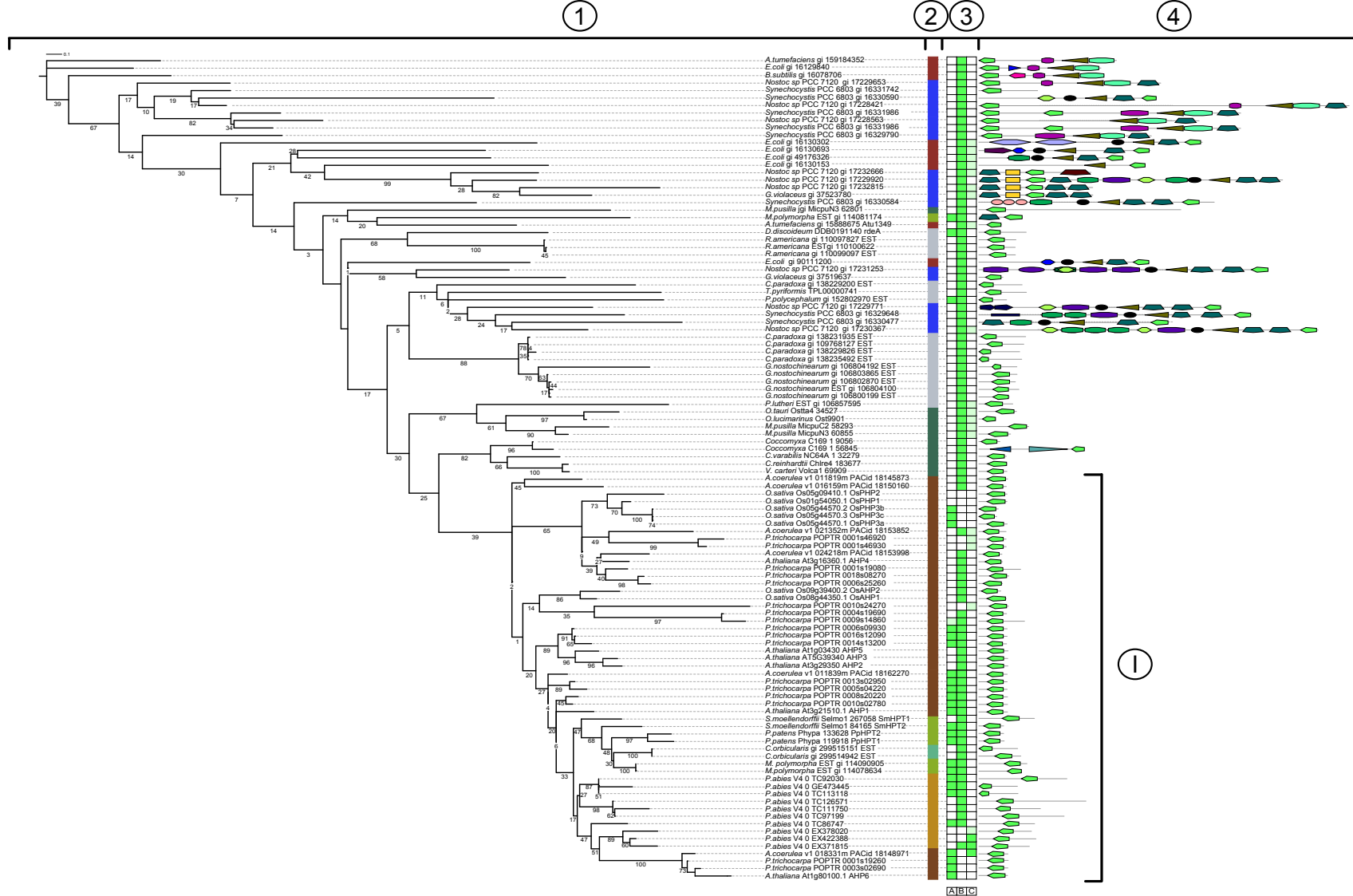
Color stripe:



Protein domains:



Supplemental Figure S2. Phylogenetic sub-trees for different domains of hybrid histidine kinases. (A) Schematic overview of the domain composition of a cytokinin receptor. For each sequences set of the (B) CHASE, (C) HisK, (D) HATPase or (E) response regulator domains were aligned and a maximum likelihood tree was calculated. In these trees clades were colored according to known plant gene functions associated with the respective protein. Sub-trees with the relevant clades are shown. For coloring the clades, the least internal node of a colored clade was defined by: the node where only plant genes derived from, no other sequences are part of the colored clades or where two functional annotations met.



1) Maximum likelihood tree of the HPT-domain.

2) Colors for grouping the organism lineages:

Angiosperm Gymnosperm Lycophyte, Bryophyte Charophyte Chlorophyte Cyanobacteria Bacteria others

3) Conservation of the canonical histidine residue and adjacent amino acids in the HPT-domain:

Position A, (minus 2 aa from conserved, canonical H)

Position B, (conserved, canonical H)

Position C, (plus 1 aa from conserved, canonical H)

Position minus 1 aa and plus 2 aa did not reveal any conclusive information about possible phosphorylation status.

Dark green shows canonical Histidine, light green marks conserved aa group, white boxes do not show any conservation related to the canonical histidine.

4) Domain architecture of the whole protein:

HisKA CHASE SBP_bac_3 GAF Response_reg P2 Trans_reg_C CheY-
CHASE3 CHASE3 HAMP PAS_3 CheW GGDEF HNOBA Pribosyltran OMPDecase
MHYT PAS PAS H- Hpt DUF2222 HATPase_c

1) Potentially cytokinin related proteins

