Running head: Phylogenetics identifies new cytokinin receptor subfamily

Corresponding Author:

Dr. Alexander Heyl
Freie Universität Berlin
Institute of Biology/Applied Genetics
Dahlem Centre of Plant Sciences
Albrecht-Thaer-Weg 6
14195 Berlin
Germany
Tel: ++49 (0)30 838 56550
Fax: ++ 49 (0)30 838 54345
heyl@zedat.fu-berlin.de

Research area: Signaling and Response
Secondary research area: Genes, Development and Evolution
Title:

A new subfamily of putative cytokinin receptors is revealed by an analysis of the evolution of the two-component signaling system of plants

Nijuscha Gruhn\textsuperscript{a}, Mhyeddeen Halawa\textsuperscript{a}, Berend Snel\textsuperscript{b,c}, Michael F. Seidl\textsuperscript{b,c} and Alexander Heyl\textsuperscript{a}.

\textsuperscript{a}Institute of Biology/Appplied Genetics, Dahlem Centre of Plant Sciences, Freie Universität Berlin, Germany

\textsuperscript{b}Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

\textsuperscript{c}Centre for BioSystems Genomics, P.O. Box 98, 6700 AB Wageningen, The Netherlands

Summary:

Phylogenetic analysis of members of the two component signaling system identifies a new subfamily of putative cytokinin receptors.
Footnotes:

Financial support

We are grateful for the financial support of the Volkswagen Foundation to N.G. (Az I/83 477) and of the Damascus University (278) and the Dahlem Centre of Plant Sciences to M.H.. Furthermore, parts of this project were carried out within the research program of the Centre of BioSystems Genomics (CBSG), which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

Current Address of M.F. Seidl: Laboratory of Phytopathology, Wageningen University, P.O. Box 8025 6700 EE Wageningen, The Netherlands
Abstract

The two-component signaling system (TCS) - 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 kinase receptor and the signal is transduced by a multi-step phospho-relay system of histidine phosphotransfer proteins (HPT) and different classes of response regulators (RR). To shed light on the origin and evolution of TCS 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 novel 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. HTPs of charophyceae seemed to be more closely related to those of land plants than to other groups of green algae. Further 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 evolutionary of a different origin than *bona fide* response regulators.
Introduction

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 multi-cellularity or from an aquatic to a terrestrial habitat (Rensing et al., 2008; Prochnik et al., 2010; Cock and Coelho, 2011). Implicit in these adaptations are the evolution of complex developmental programs. The execution of those programs is regulated by a multi-layered interplay of different plant hormones (Jaillais and Chory, 2010; Vanstraelen and Benkova, 2012; El-Showk et al., 2013).

One class of phytohormones is a group of $N^\beta$-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 system (TCS). However, bacteria do not respond to cytokinin (Spichal, 2012). In its simplest form the TCS consists of a receptor histidine kinase, which autophosphorylates upon signal perception, and a response regulator (RR), which mediates the output after being activated by phosphorylation of a canonical Asp in the response regulator domain. Cytokinin receptors are hybrid histidine kinases, as they contain both a histidine kinase and a response regulator domain. The cytokinin ligand is bound via the cyclase/histidine kinase associated sensory extracellular (CHASE) domain (Anantharaman and Aravind, 2001; Mougel and Zhulin, 2001; Heyl et al., 2007), and this binding is thought to trigger a conformational change leading to the autophosphorylation of the receptor (Miwa et al., 2007; Hothorn et al., 2011). After an intra-molecular transfer from the histidine kinase to the response regulator domain of the receptor, the phosphate is transferred to histidine phosphotransfer proteins (HPTs). These proteins shuttle continuously between the cytoplasm and the nucleus (Punwani et al., 2010). In the nucleus they can phosphorylate the type-B response regulators (RRB). The RRBs are transcription factors containing a response regulator domain and a Myb-related DNA binding domain, which allows them to bind to their target DNA sequences (Sakai et al., 2000; Hosoda et al., 2002). One group of their target genes are the type-A response regulators (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 in other plants (Hellmann et al., 2010). Previous studies identified an additional group of response regulators, the “pseudo”- response regulators (PRRs), members of which were shown to have a role in the regulation of circadian rhythm and the type-C response regulators (RRC) for which a clear biological function has yet to be determined (Mizuno, 2004).

The ability of plants to use cytokinin as a phytohormone represents an evolutionary novelty (Gruhn and Heyl, 2013). This raises the question 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 this question 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 new subfamily of cytokinin receptors found only in early diverging land plants. While their domain architecture is similar to 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 demonstrated the ability of two members of this subfamily to bind cytokinin and to translate the binding of different types of cytokinins into a cellular signal. Furthermore we found first hints for the presence of cytokinin receptors in the charophyceae algal species Spirogyra pratensis, the group of green algae giving rise to land plants. In addition, the analysis revealed the presence of RRBs, which display diverse domain architecture. 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 new 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 sub-clades (Fig. 1). While most of the land plant CHASE-domains clustered similarly to what was published previously (Pils and Heyl, 2009), a new clade containing eight sequences from *Physcomitrella patens* and one from *Marchantia polymorpha* previously not associated with cytokinin signaling emerged. In addition, a third clade contained CHASE domains from receptors of cyanobacteria, *Dictyostelium*, 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, HisK, HATPase, and RR domain) (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 which were shown to be important for its structure and the cytokinin binding (Heyl et al., 2007; Hothorn et al., 2011), a high level of conservation was found among the “classical” cytokinin receptors, while members of the novel subfamily had only a very low level of conservation compared to the CHASE domain of CRE1/AHK4 (Fig. 1B). To investigate the new 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 from 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 indicates that most members of this new family are indeed expressed.

In order to test experimentally whether members of this clade can function as cytokinin receptors, we selected two proteins to serve as examples. MpCHK1 (*Marchantia polymorpha* CHASE domain containing Histidine Kinase receptor) is the only detected putative cytokinin receptor from *Marchantia polymorpha*, the earliest
diverging land plant in this study. For the three new PpCHKS for which we found EST evidence for their expression in both analyzed moss species, one (PpCHK4) was randomly chosen for further analysis. Both proteins, PpCHK4 and MpCHK1, were expressed in E.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 was 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 that detected for the AHK4 (T301I) (Fig. 2). However, also due to 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 new receptors (Mizuno and Yamashino, 2010). The two potential cytokinin receptors activated the reporter gene specifically in response to various cytokinins in a dose-depended manner. Trans-Zeatin treatment resulted in the strongest activation of the reporter gene for all three functional receptors. Interestingly, while cis-Zeatin caused the weakest response of all tested cytokinins in AHK4 of Arabidopsis, it triggered a stronger activation than e.g. BA in the moss receptor PpHCK4. This 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. This data demonstrate that at least two members of the new clade of putative cytokinin receptors have the biochemical characteristics consistent with their domain architecture and are thus 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 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. 1), the HATPase, or the HisK domain was used to generate a phylogenetic tree, we observed similar relationships among the dataset (Supplemental Fig. 2). Sequences that clustered together for one domain were also found in the same sub-clades for the other domains. Clearly recognizable subfamilies of histidine kinases were those grouping with the osmosensor AHK1 (*A. thaliana*) (Tran et al., 2007), the sensor kinases CKI1 (*A. thaliana*) (Kakimoto, 1996) and AHK5 (*A. thaliana*) (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 response regulator tree the clade representing the histidine kinases with a CHASE domain contained four EST sequences that have only a RR domain (Supplemental Fig. 1). These sequences originated from the gymnosperm *Picea abies* and from the charyophyte alga *Spirogyra pratensis* and might be parts of complete cytokinin receptors. This hypothesis is further supported by the clustering of a HATPase domain from an *Spirogyra pratensis* EST in the clade of the HATPase domain of cytokinin receptors (Supplemental Fig. 2). Thus the incorporation of EST data in the analysis provided the first hints at an origin of the cytokinin receptors in charophyceae algae, outside of the land plants.

Evolution of the histidine phosphotransfer proteins recapitulates plant evolution

The analysis of the histidine phosphotransfer proteins (HPTs) 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 (Maximum Likelihood or Bayesian interference). The analysis of the HPTs showed that members of this protein family from bacteria were distinct from those 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 a single exception all HPTs from the
green lineage did not contain any additional domains, such as GAF or CheY domains (Fig. 4). For Arabidopsis it was shown that one of the HPTs (AHP6) does not contain the canonical histidine residue and thus cannot be phosphorylated. However, it functions as a negative regulator of cytokinin signaling (Mähönen et al., 2006). An analysis for the presence of the conserved histidine residue in the HPT domains revealed that only angiosperms and gymnosperms contained such non-canonical HPTs, as no proteins carrying such a mutation were found in any of the other clades (Fig. 4 and Supplemental Fig. 3).

Response regulators show distinct origins

The phylogenetic analysis of the response regulator domain revealed the RRAs and the RRBs to be closely related to each other, while the RRCs grouped to the clade of the RR domains of histidine kinases (Supplemental Fig.1). 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). While 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 two component signaling (Horák et al., 2008). All RR domains from the dataset were analyzed in one phylogenetic tree (Supplemental Fig. 1). This tree showed that the RR domains of the RRCs are most closely related to those of histidine kinases from plants than to the RRAs or RRBs. This 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 a HATPase domain in two members of this clade (Fig. 4). While 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 sub-clades (Fig. 6). Although the individual branches differed depending on the phylogenetic method (ML or Bayesian interference), the three major subclades (RRA, RRB, PRR) were supported by both methods (Supplemental Fig. 4). All known RRAs were found in one monophyletic group with sequences exclusively from land plants. While in almost all of the sequences the canonical Asp was conserved, in three cases, intriguingly restricted to spruce, this critical residue was substituted. The other sub-clade was further subdivided into the RRB branch and a branch containing the so called “pseudo-response regulators” (Makino et al., 2000). PRRs often contain additional domains apart from the RR- and the Myb-domains, such as a CTT (CONSTANS, CO-like, and TOC1) domain, and function in circadian rhythm (Mizuno, 2004). Our phylogenetic analysis shows a clear separation between the RRBs and the PRRs. This was also reflected in the domain structure, as most PRRs contain also a CCT or a WD40 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 might therefore 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 from algae and basal land plants, shared the conservation of the canonical Asp of the RR domain with the *bona fide* response regulators 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 are central questions in biology. 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, has
become possible. Taken 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 to use 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, e.g. 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. 2). 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 due to their probability of occurrence but also are in agreement with the available biological data.

Cytokinin perception via 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 new subfamily of cytokinin receptors, which were shown to bind the phytohormone and to translate this binding into a cellular response. At this point we can only speculate about a possible biological function of members of this new 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 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, this could be either due to our limited access to the genome of this liverwort, as we obtained only BLAST results from the genomic sequence using AHK4 of Arabidopsis as a query or that there is only one cytokinin receptor in this species. Thus there might be more cytokinin receptors in the Marchantia genome, which have not yet been identified. Only the sequencing of more genomes from charophycean 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 as recent analysis found high sequence conservation in the amino acids on both sides of the CHASE domain (Steklov et al, 2013). A second surprising result of this analysis was the detection of two ESTs in the response regulator tree and one EST in the HATPase tree of the charophyte alga Spirogyra pratensis that clustered with those of the cytokinin receptors. These results suggest that those ESTs might be part of cytokinin receptors or an 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 histidine kinases rather than with the other RRs shown before (Kiba et al, 2004; Schaller et al, 2008). Remarkably, some RRCs and also some hybrid histidine kinases share a phosphatase activity and might thus work as negative regulators of two-component signaling pathways (Kiba et al., 2004; Mähönen et al., 2006; Horák et al., 2008). Further evidence for RRCs as degenerated hybrid histidine 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. In this context it is worthwhile mentioned that the overexpression of one of the RRC from Arabidopsis, ARR22, leads to a similar phenotype as the wol mutation of the cytokinin receptor mutant AHK4 (Kiba et al., 2004). But 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 large protein family.

Phylogenetic analysis of histidine phosphotransfer 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 multi-domain proteins, e.g. receptors, while in unicellular eukaryotes and plants the HPT domain is the only domain of a short protein. This 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 response regulators. Thus the need arises for HPTs which can easily shuttle between the two compartments, as it was shown for Arabidopsis (Punwani et al., 2010). In contrast in bacteria, which lack the nuclear membrane, response regulators can interact directly with the histidine kinase receptors. Thus there is no need for membrane crossing HPTs.

A second important result from the phylogenetic analysis of the HPT sequences concerns the canonical histidine residue. While this critical amino acid is a prerequisite for a functioning in the His-to-Asp phosphorelay, it has been shown that also family members lacking it can play a role in cytokinin signaling (Mähönen et al., 2006). In our analysis non-canonical 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.

Type-A and type-B response regulators share a common origin

The results of the phylogenetic analysis revealed a different picture for the non-receptor types of RRs. The RRBs are plant specific and were found throughout the plant kingdom. This together with the fact that many RRBs were missing either the Myb-related DNA binding domain and/or the canonical Asp residue of the response regulator domain might indicate that they represent an evolutionarily older protein family as compared to the RRAs or that the selective force acting on them is lower than on RRAs. However, even the non-canonical 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 sub-clades. Interestingly it was shown that some PRRs have similar function as regulators of the circadian rhythm in Chlamydomonas and the land plant Arabidopsis (Mizuno, 2004; Matsuo et al., 2008).
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, while 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 a last common ancestor of RRAs and RRBs happened after the split of the chlorophyceae and the charophyceae.

Conclusions

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 not just at the evolutionary history of a single gene, but the whole associated pathway. In addition, the results of our analysis, especially the identification of a new subfamily of cytokinin receptors, vindicate the use of a domain-based approach to identify members of protein families, especially when the data includes ESTs.

Taken 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 (Frebort 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

As 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 eukaryotes, 12 chlorophyte algae, three charophyte algae and 8 land plant species (Supplementary Tab. S2). Available nucleotide sequences (EST) were translated into all six frames by 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) (Supplementary Tab. S3). The respective proteomes were analyzed for the existence of the relevant domain by HMM search of the HMMER3 Package (HMMER 3.0) (Finn et al., 2011) using the PFAM-A model available on the PFAM website (Supplementary Tab. 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 (HMMER 2.4i). The HMMER2 domain models were based on the seed alignments initially used to generate the HMMER3 models. The resulting models were subsequently calibrated using hmmcalibrate (HMMER2 package). Domains were extracted according to the HMM2 result from the protein sequence and aligned. For multiple sequence alignments (MSA) the MAFFT package (MAFFT v6.846b;(Katoh et al., 2009)) using the local pair-wise aliminent option was applied. For tree construction and bootstrapping RAxML package (RAxML v7.0.4; (Stamatakis, 2006)) was employed running under fast optimization method and search for best-scoring maximum likelihood (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/index.php) with the
‘Whelan And Goldman’ amino acid substitution matrix and the gamma shaped rate variation with a proportion of invariable sites was estimated using eight rate categories. MCMC was run with five independent runs of four chains, T=0.05, a sample frequency of 10 for 5000000 (RR and CHASE domain) or 10000000 (HPT) generations. At this point the average standard deviation of split frequency, a measure of convergence, was well below 0.02 for each analysis and the runs were stopped. The first 25% 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, critical for the function of the investigated proteins, were compared. The characteristic amino acid of the protein in the tree founding alignment was compared to published, characterized residues in known proteins. Since *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* histidine kinase AHK4, AHK4 (T301I) and the putative cytokinin receptors *Physcomitrella patens* (Pp) receptor PpCHK4 and the *Marchantia polymorpha* (Mp) receptor MpCHK1 were cloned into the pDEST15 vector (Invitrogen, Karlsruhe, Germany) and expressed using the *E. coli* strain KML002 ((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 (Prague, Czech Republic). The cDNAs for MpCHK1 and PpCHK4 were obtained by gene synthesis (Genscript,
Piscataway, NJ, USA). The protein expression was confirmed by Western blot (Heyl et al., 2007).

**Semi-quantitative complementation assay**

The functionality of the putative new cytokinin receptors was tested by a semi-quantitative complementation assay as described before (Mizuno and Yamashino, 2010). The *E. coli* stain, KMI002 was transformed with the vector pDEST15 (Invitrogen, Karlsruhe, Germany) expressing the *Arabidopsis* histidine kinase AHK4, AHK4 (T301I), the putative *Physcomitrella patens* receptor PpCHK4 or the putative *Marchantia polymorpha* receptor MpCHK1, respectively.

**Acknowledgements**

We want to thank Sandra K. Floyd and John Bowman from the Monash University, Melbourne, 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 under Contract No. DE-AC02-05CH1123, for sharing the results of BLAST searches of the so far unpublished *M. polymorpha* genome with us.

**References**


Figure Legends

Figure 1. Phylogenetic tree of CHASE domain containing histidine kinases reveals a new subfamily of plant cytokinin receptors. (A) The Cyclase/ Histidine kinase Associated Sensory Extracellular (CHASE) domain sequences from all organisms under investigation were aligned and a phylogenetic tree using a Maximum Likelihood and Bayesian methods were inferred (for details see material and methods section). Since tree topology was identical with both methods, bootstrap values and posterior probabilities are depicted on the Maximum Likelihood 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 filled circle and the lack of EST evidence by an open circle. (B) Left: Maximum likelihood tree of the CHASE domain, Middle: Conservation of amino acids important for 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 (For details see Supplemental Tab. S1). Dark blue shows identity to AHK4, light blue marks conservative substitution, white boxes symbolizes no conservation related to the respective position in the CHASE sequence of AHK4. Right: Domain architecture of the whole protein. For details see the materials and methods section.

**Figure 2.** Two members of the novel subfamily of putative cytokinin receptors bind cytokinin. (A) *In vitro* binding of trans-[2-3H]Zeatin to AHK4; PpCHK4; MpCHK1 and AHK4 (T301I) proteins, respectively, overexpressed in *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 and GST-antibody. The arrow highlights the band for AHK4.

**Figure 3.** Experimental evidence for a function as cytokinin receptors for two members of the novel subfamily. AHK4, PpCHK4; MpCHK1 and AHK4 (T301I) were tested in a bacterial complementation assay with different cytokinin and adenine concentrations. Abbreviations (tZ, *trans*-Zeatin; BA, Benzyladenine; cZ, *cis*-Zeatin; IP, Isopentenyl; Ade, Adenine)

**Figure 4.** Phylogenetic tree for the different histidine phosphotransfer proteins (HPT) investigated in this study. Left: Sequences of the histidine phosphotransfer domain were aligned and a phylogenetic tree using Maximum Likelihood or Bayesian methods was inferred (black lines). Both trees show identical topology, therefore bootstrap and posterior probabilities are depicted on the tree. For protein identifiers please refer to Supplemental Fig. S3. The clade containing proteins involved in cytokinin signaling is highlighted with blue lines. Canonical His: Green indicates conserved His residue on relevant position, white boxes do not show any conservation. Right: Domain architecture of the whole respective protein. For details see materials and methods section.
**Figure 5.** Phylogenetic tree for the type-C response regulators. Left: The sequences of the phylogenetic sub-tree of the type-C response regulator clade identified previously (Supplemental Fig. S1) were aligned and a maximum likelihood tree was calculated. Canonical Asp: Green indicates conserved Asp residue on relevant position, white boxes do not show any conservation. Right: Domain architecture of the whole respective protein. For details see materials and methods section.

**Figure 6.** Phylogenetic relationship of type-A, type-B and pseudo response regulators. Left: Sequences of the response regulator (RR) domain were aligned and a maximum likelihood tree was calculated. Depicted is a subtree (for protein identifiers please refer to Supplemental Fig. S1) containing the A-type (red lines), B-type (blue lines) and pseudo RR (gray lines). Canonical Asp: Dark blue shows conservation of canonical Asp, light blue marks conservative substitution, white boxes symbolizes no conservation. Right: Domain architecture of the whole respective protein. For details see the materials and methods section.

**Supplemental Figures:**

**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.

**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.

**Supplemental Figure S3.** Phylogenetic tree for the different histidine phosphotransfer proteins (HPT) investigated in this study. This represents the same phylogenetic tree as shown in Fig. 3, but with the sequence identifier added.

**Supplemental Figure S4.** Phylogenetic relationship of type-A, type-B and pseudo response regulators. The phylogenetic tree for the subclades 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.

**Supplemental Figure S5.** Phylogenetic relationship of type-A, type-B and pseudo response regulators (complete identifiers). This represents the same phylogenetic tree as shown in Fig. 4, but with the sequence identifier added.

**Supplemental Tables:**

**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 domain investigated.
**Figure 1.** Phylogenetic tree of CHASE domain containing histidine kinases reveals a new subfamily of plant cytokinin receptors. (A) The Cyclase/ Histidine kinase Associated Sensory Extracellular (CHASE) domain sequences from all organisms under investigation were aligned and a phylogenetic tree using a Maximum Likelihood and Bayesian methods were inferred (for details see materials and methods section). Since tree topology was identical in both methods, bootstrap values and posterior probabilities are depicted on the maximum likelihood tree. The newly identified subfamily of plant cytokinin receptors is indicated with red lines and the clad with classical cytokinin receptors is marked by blue lines. EST evidence is symbolized by a filled circle and the lack of EST evidence by an open circle. (B) The phylogenetic tree is the same as in Fig. 1 A Left: Maximum likelihood tree of the CHASE domain. Middle: Conservation of amino acids important for 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 (For details see Supplemental Tab. S1). Dark blue shows identity to AHK4, light blue marks conservative substitution, white boxes symbolizes no conservation related to the respective position in the CHASE sequence of AHK4. Right: Domain architecture of the whole protein. For details see the Materials and Methods section.
Figure 2. Two members of the novel subfamily of putative cytokinin receptors bind cytokinin. (A) In vitro binding of trans-[2-3H]zeatin to AHK4, PpCHK4, MpCHK1 and AHK4 (T301I) proteins, respectively, overexpressed in E. coli strain KM002. 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 and GST antibody. The arrow highlights the band for AHK4.
Figure 3: Experimental evidence for a function as cytokinin receptors for two members of the novel subfamily, AHK4, PpCHK4, MpCHK1 and AHK4 (T3011) were tested in a bacterial complementation assay with different cytokinin and Adenine concentrations. Abbreviations (I2, trans-Zeatin; BA, Benzyladenine; cZ, cis-Zeatin; IP, Isopentenyl; Ade, Adenine)
Figure 4. Phylogenetic tree for the different histidine phosphotransfer proteins (HPT) investigated in this study. Left: Sequences of the histidine phosphotransfer domain were aligned and a phylogenetic tree using Maximum Likelihood or Bayesian methods was inferred (black lines). Both trees show identical topology, therefore bootstrap and posterior probabilities are depicted on the tree. For protein identifiers please refer to Supplemental Fig. S3. The clade containing proteins involved in cytokinin signaling is highlighted with blue lines. Canonical His: Green indicates conserved His residue on relevant position, white boxes do not show any conservation. Right: Domain architecture of the whole respective protein. For details see materials and methods section.
Figure 5. Phylogenetic tree for the type-C response regulators. Left: The sequences of the phylogenetic sub-tree of the type-C response regulator clade identified previously (Supplemental Fig. S1) were aligned and a maximum likelihood tree was calculated. Canonical Asp: Green indicates conserved Asp residue on relevant position, white boxes do not show any conservation. Right: Domain architecture of the whole respective protein. For details see materials and methods section.
Figure 6. Phylogenetic relationship of type-A, type-B and pseudo response regulators. Left: Sequences of the response regulator (RR) domain were aligned and a maximum likelihood tree was calculated. Depicted is a subtree (for protein identifiers please refer to Supplemental Fig. S5) from the global response regulator tree (Supplemental Figure S1) containing the A-type (red lines), B-type (blue lines) and pseudo RR (gray lines). Canonical Asp: Dark blue shows conservation of canonical Asp, light blue marks conservative substitution, white boxes symbolizes no conservation. Right: Domain architecture of the whole respective protein. For details see the materials and methods section.