Evolution and phyletic distribution of two-component signal transduction systems
Kristin Wuichet¹, Brian J Cantwell¹ and Igor B Zhulin¹,²

Two-component signal transduction systems are abundant in prokaryotes. They enable cells to adjust multiple cellular functions in response to changing environmental conditions. These systems are also found, although in much smaller numbers, in lower eukaryotes and plants, where they appear to control a few very specific functions. Two-component systems have evolved in Bacteria from much simpler one-component systems bringing about the benefit of extracellular versus intracellular sensing. We review reports establishing the origins of two-component systems and documenting their occurrence in major lineages of Life.

Introduction
Signal transduction systems regulate cellular processes in all living organisms. The evolutionary success of prokaryotes is dependent on the ability of these unicellular organisms to rapidly sense and respond to changes inside and outside their cell. Similarly, eukaryotes also must detect and transmit both internal and external (e.g. cell-to-cell) signals to adjust functions controlled by hormones, cytokines, mediators, and so on. Two-component regulatory systems (TCS) were the first major mode of signal transduction identified in bacteria [1]. Since the time of their discovery, many TCS have been identified and studied, primarily in prokaryotes [2**]. Advent of genomics led to massive identification of TCSs in microbial genomes [3,4*]. TCS have also been identified in eukaryotes [5]. It appears that their distribution in eukaryotes, specifically lower eukaryotes and plants, is significantly less abundant [2**,6]. Many bacterial TCSs have been shown to play a role in virulence [7,8], and the absence of the systems in mammals furthers interest in their study as potential drug targets [9]. Understanding the evolution of TCSs will expand knowledge about these thoroughly studied systems to derive both system-specific and universal principles of signal transduction. There have been several attempts to address the question of TCS origins and diversification [10**,11**,12,13]. Here, we summarize both advances and shortcomings of our current understanding in this important area.

System design: one component, two components...
As their name suggests, two-component signal transduction systems are composed of two dedicated proteins, a sensor and a regulator [14**], although some systems contain additional auxiliary components. Interestingly, there are other signal transduction systems that appear to be either simpler or more complex in their overall design than classical TCSs. The majority of sensing and signal transduction in prokaryotes appears to be carried out by so-called one-component systems (OCSs), single proteins that combine properties of both a sensor and a regulator [10**]. Usually, these properties reside in two distinct domains—sensory and regulatory (Figure 1), although they can be combined in a single domain or distributed between multiple domains within a single protein [4*,10**]. The relatedness between OCSs and TCSs lies in the fact that they share a repertoire of sensory and regulatory domains [10**].

The prototypical TCS employs four principal domains: sensory (also called input) and transmitter that reside in a sensor histidine kinase (HK) and receiver and regulatory (also called output) that reside in a response regulator (RR) (Figure 1). The transmitter and receiver communicate via phosphorylation [2**]. Many TCSs stray from this paradigm by employing multiple sensory domains (both extracellular and cytoplasmic) [15,16], and additional transmitter and regulatory domains, as well as intermediate components between the kinase and response regulator that usually comprise extended phosphorelays [2**,17]. Sensor kinases with C-terminal receiver domains are referred to as hybrid histidine kinases (HHK) [2**]. Similarly there are also hybrid response regulators (HRR) with C-terminal HK modules rather than typical output domains. HHKs and HRRs are ubiquitous, but still far outnumbered by HKs and RR with standard architectures (Table 1). In addition to sensor, transmitter (kinase), receiver, and regulatory modules,
histidine phosphotransfer (HPT) domains are commonly associated with TCSs. They often can be found fused to HHKs or as independent proteins to facilitate the phosphotransfer process [2,17]. HPT domains can both receive phosphoryl groups from and transmit them to receiver domains [18]. As seen with HPT domains, receiver domains can also exist as independent proteins that lack additional output domains, which are members of extended phosphorelays, allosteric regulators, or other roles [19]. The modular nature of these domains leads to a variety of architectures (Figure 2), resulting in diverse phosphorelay pathways.

At the apex of TCS complexity is the bacterial chemotaxis system, a specialized TCS that utilizes a dedicated chemoreceptor (sensor), a sensor-less HK (CheA protein), an RR protein that lacks a usual output domain (CheY protein), and a number of specialized auxiliary components (Figure 1). Genomic modularity of this system is considered elsewhere [20] and is beyond the scope of this review.

Phyletic distribution: abundance in prokaryotes and instances in eukaryotes

A genomic survey shows that TCSs are found in all three superkingdoms: Bacteria, Archaea, and Eukarya (Table 1). TCS components are present in 864 of 899 available completely sequenced bacterial genomes including all 21 phyla represented (Table 1). The only bacterial species lacking TCSs are pathogens (e.g., Mycoplasma species) and endosymbionts (e.g., Amoebophilus species) with severely reduced genomes (mistdb.com).
Previously TCSs were considered absent from the Tenericutes phylum [10,11], but the newly available genome of *Acholeplasma laidlawii* revealed multiple HKs and RR (mistdb.com). In contrast to Bacteria, TCSs are present only in 50% of genomes from Archaea. They appear to be entirely absent in Crenarchaeota, Korarchaeota, and Nanoarchaeota phyla. TCSs are found in 33 of the 42 genomes representing Euryarchaeota, and the one representative of Thaumarchaeota, *Nitrosopumilus maritimus* SCM1. Less than 30% of eukaryotic genomes encode TCSs (Table 1). Queries of the SMART database identify TCSs in fungal, plant, and slime mold lineages, but they are absent from metazoan and protist lineages, as suggested by an earlier genome survey [3].

| Genome surveys have been carried out using the MiST2 database [4] for Bacteria and Archaea and the genomic mode of the SMART 6 database [56] for Eukarya. CheA are not included in this survey. Domain architecture queries for HATPase_c, REC, and HPT domains were used to identify putative TCS components in SMART, which were then manually classified. BLAST queries using TCS sequences identified in Metazoan genomes in SMART support that they are the result of bacterial contamination and were subsequently excluded. TCS components are automatically classified in MiST2. In MiST2 the 'other' category of TCS components predominantly consists of stand-alone HPT proteins, but rarely includes proteins with architectures that do not fit the standard rules for TCS classification. HPT proteins are absent from Archaea. Bacteria and Archaea taxonomic groups are comprised of phyla as defined in the MiST database [4]. Given the fewer representatives of eukaryotic phyla and their comparatively recent emergence in comparison to prokaryotic phyla, Eukarya taxonomic groups correspond to the more inclusive fungal, metazoan, plant, protist, and slime mold lineages.

The distribution of HHKs and HRRs in Bacteria, Archaea, and Eukarya reveal differences in their predominant TCS signal transduction mechanisms. HHKs make up less than 20% of TCS kinases in bacteria, and they are found in only 16 of the 21 phyla (Table 1). Surprisingly, more than 90% of eukaryotic TCS kinases are HHKs, and they are present in all three lineages that have TCSs. Only 1% of the TCS kinases in Archaea are HHKs, and they are found exclusively in Euryarchaeota. HRRs are entirely absent from Eukarya and comprise

| Table 1: Genomic and phyletic distribution of TCS components in three domains of Life |
|-----------------|-----------|-----------|
| Genome occurrence | Bacteria | Archaea | Eukarya |
| Genomes | 899 | 67 | 55 |
| HK | 857 | 30 | 1 |
| RR | 862 | 34 | 15 |
| HHK | 561 | 5 | 15 |
| HRR | 258 | 17 | 0 |
| Other/HPT | 220 | 2 | 15 |

| Taxonomic distribution | Bacteria | Archaea | Eukarya |
|-----------------|-----------|-----------|
| Taxonomic groups | 21 | 5 | 5 |
| HK | 21 | 3 | 1 |
| RR | 21 | 2 | 3 |
| HHK | 16 | 1 | 3 |
| HRR | 15 | 1 | 0 |
| Other/HPT | 11 | 1 | 3 |

| Total TCS counts | Bacteria | Archaea | Eukarya |
|-----------------|-----------|-----------|
| HK | 20,862 | 546 | 7 |
| RR | 26,962 | 304 | 80 |
| HHK | 4717 | 8 | 86 |
| HRR | 923 | 142 | 0 |
| Other/HPT | 329 | 4 | 18 |

The distribution of HHKs and HRRs in Bacteria, Archaea, and Eukarya reveals differences in their predominant TCS signal transduction mechanisms. HHKs make up less than 20% of TCS kinases in bacteria, and they are found in only 16 of the 21 phyla (Table 1). Surprisingly, more than 90% of eukaryotic TCS kinases are HHKs, and they are present in all three lineages that have TCSs. Only 1% of the TCS kinases in Archaea are HHKs, and they are found exclusively in Euryarchaeota. HRRs are entirely absent from Eukarya and comprise

---

only 3% of the RRs in Bacteria; however, they make up more than 30% of the RRs in Archaea. While multiple TCSs employing HHKs have been studied in both Bacteria and Eukarya [6,17], there is an astonishing lack of experimental information on HRRs. This is particularly relevant to Archaea (see Table 1).

Given the vast numbers of TCSs in Bacteria, which have been shown to control a variety of cellular processes [21], it is nearly impossible to review them in a traditional way. Spectacularly, there is little to no information on TCSs in Archaea, but experimental studies have revealed interesting aspects of TCSs in eukaryotes. Unicellular fungi typically encode a small number of histidine kinases [22,23], while the genome sequences of various filamentous fungi have revealed between 11 and 21 histidine kinases [23,24,25,26]. All of the kinases identified in fungi are hybrid in and in all cases, these organisms encode a single HPT protein [23,26]. The fungi typically encode two RRs, but some of the filamentous fungi possess a third response regulator as well [26]. The slime mold Dictyostelium discoideum encodes 21 two-component proteins. These include 14 HHKs and 6 RRs [27]. Similarly to all fungi [23,26], D. discoideum encodes only a single HPT protein, indicating that the sensor kinases may integrate multiple signals through a single phosphorylation [27]. Compared to the lower eukaryotes, plants encode a greater number and diversity of phosphorelay proteins. Analysis of the Arabidopsis thaliana and Oryza sativa genomes identified 14–16 HKs, 5 HPT proteins, and 32 RRs in each [28,29]. In contrast to fungi and D. discoideum, HHKs represent approximately half of the plant TCSs rather than all, but the importance of HPT proteins in these systems is evident by their expansion. This is in stark contrast with prokaryotes that possess HHKs, fewer than 40% of which encode an HPT protein in their genomes (Table 1), which is consistent with the results of a previous survey [6].

Evidence for Bacterial origins and horizontal gene transfer

There are multiple evolutionary forces driving the emergence and expansion of TCS diversity that can be visualized using computational approaches. TCS components are often classified on the basis of their domain similarity and RRs are integral components of the HK. D. discoideum encodes only a single HPT protein, indicating that the sensor kinases may integrate multiple signals through a single phosphorylation [27]. Compared to the lower eukaryotes, plants encode a greater number and diversity of phosphorelay proteins. Analysis of the Arabidopsis thaliana and Oryza sativa genomes identified 14–16 HKs, 5 HPT proteins, and 32 RRs in each [28,29]. In contrast to fungi and D. discoideum, HHKs represent approximately half of the plant TCSs rather than all, but the importance of HPT proteins in these systems is evident by their expansion. This is in stark contrast with prokaryotes that possess HHKs, fewer than 40% of which encode an HPT protein in their genomes (Table 1), which is consistent with the results of a previous survey [6].

The co-evolutionary relationship between cognate HKs and RRs is evident in the congruent phylogenetic trees built from multiple sequence alignments of these proteins [11]. Cognate HK and RR proteins are often encoded adjacent to each other in genomes, which further supports their co-evolutionary relationship [4,10]. Large-scale efforts to deduce interacting HKs and RRs in multiple genomes have relied on genome context comparison to Archaea [10]. However, many systems utilize orphan HKs and RRs that are not encoded adjacent to their cognate partners [4], and recent experimental studies support their importance in multiple organisms [33–36]. Furthermore, genome context and phylogenetic methods do not always agree when predicting cognate HKs and RRs [11]. Domain architecture-based classification can be totally misleading when used without phylogenetic support. This is illustrated by stunning differences in the domain architectures of orthologous HKs that control nitrogen fixation (FixL) in Rhizobia [37–39].

TCS evolution in eukaryotes

Substantial phylogenetic evidence in light of the known pattern of distribution of eukaryotic TCSs led to a conclusion that the TCSs in eukaryotes represent multiple independent horizontal gene transfer events from bacteria to eukaryotes that occurred at a time well after the split into the three superkingdoms [11]. Many genes in plant nuclear genomes are thought to have been transferred from the chromosomes of cyanobacterial symbionts during the evolution of the chloroplast, with as much as 18% of the protein-coding genes of A. thaliana inherited from cyanobacteria [40]. The CSK protein of A. thaliana is a HK, which is encoded by a nuclear gene and targeted to the chloroplast. The cognate RR of the CSK is plastid encoded, and phylogenetic analysis of this family of proteins identified the cyanobacterial HK, Hik2, as a common ancestor [41].

TCSs in eukaryotes have undergone a number of evolutionary innovations distinct from those seen commonly in bacterial systems. One such feature is the integration of TCS signaling pathways with other signaling systems typical of eukaryotes. The output of the fungal TCSs feeds into a mitogen-activated protein kinase (MAPK) pathway. This is best studied in Saccharomyces cerevisiae.
where the Sln1-Ypd1-Ssk1 phosphorelay activates the Ssk2 MAPK leading to control of genes for osmotic stress [42]. Similar pathways are found in other fungal species studied [26]. Although RRks in yeast can be regulated by phosphorylation, one has been shown to affect gene expression in a manner independent of aspartate phosphorylation that is typical of prokaryotes [43]. The RdeA-RegA phosphorelay in D. discoideum have been shown to regulate cAMP levels [44] and the DskA HK can be regulated by serine phosphorylation [45]. The A. thaliana ethylene receptors interact with the serine-threonine kinase CTR1 [46] and mediates responses to ethylene independent of histidine-aspartate phosphotransfer [47]. The plant cytokinin TCSs appear to function similarly to the common bacterial systems and directly regulate gene expression through phosphorylation of a DNA-binding RR [48]. However, even in this case there are significant differences, as cytokinin signaling seems to involve considerable crosstalk with multiple HKs and RRs communicating through a common pool of HPT proteins [49]. The phytochrome receptor proteins of plants contain a divergent HK domain that possesses serine-threonine kinase activity [50]. A family of RRs in A. thaliana is not phosphorylated [51], but some play important roles in the regulation of the circadian clock [52].

Conclusions

Comparative genomics methods prove to be effective for describing the diversity of TCS components and tracing their evolution. Such information is currently inaccessible by experimental methods alone. While domain architecture and genome context analyses can visualize the evolutionary forces effecting TCS diversity, such as domain duplication, domain/gene birth and death, gene duplication/divergence, and gene fission/fusion events [6,12,15,53,54], there are still many unanswered questions about the evolution of these systems that cannot be addressed without an exhaustive phylogenomic approach (Box 1). Special effort must be made to be comprehensive, particularly to reduce erroneous conclusions that can arise from analyses using only a single method. Similarly, analyses utilizing only a single genome [55] can only provide limited information that may not be applicable to the larger landscape of TCS evolution and function. Current studies are primarily descriptive and rely on a single methodology. Future efforts utilizing systems-level phylogenomics are required in order to properly classify and understand complete signal transduction pathways.

Box 1 Future directions and unanswered questions

- What levels of complexity can be achieved in extended phosphorelays, particularly those involving HRRs? What is the exact function of HRR?
- How can we efficiently assign orphan HKs/HHKs and RRs/HRRs to their cognate partners, especially in the light of the exponential growth of genomic data?
- How often orthogonal HKs and RRs perform different functions in different organisms? Defining orthology is the main approach to protein function prediction. How careful should we be in using orthology for function prediction in signal transduction proteins?
- Sensory specificity of HKs remains a single less-studied topic in two-component signal transduction. We should develop high-throughput computational and experimental methodologies to address this problem.
- Various high-throughput computational and experimental methods are now being used to predict targets for transcriptional regulators, including RRs. A comprehensive comparison of these methods and a unified approach must be developed in the nearest future.

References and recommended reading

Paper of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


This paper describes a continuously updated database that classifies TCS components in all prokaryotic genomes. It includes the first attempts to delineate evolutionary relationships between TCS proteins on the basis of domain orthology and also catalogs HKH and HRR proteins as distinct types of TCS components.


While individual proteins that could both sense an input and effect an output had been identified previously in experimental studies, this paper was the first to establish the dominance of OCSs as the major mode of signal transduction in prokaryotes and their placement as the progenitors of TCSs.


14. Kofoid EC, Parkinson JS: **Transmitter and receiver modules in bacterial signaling proteins.** Proc Natl Acad Sci USA 1988, 85:4961-4965. This paper uncovers the universal modularity of TCS.


19. Jenal U, Galperin MY: **Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics.** Curr Opin Microbiol 2009, 12:152-160. This review provides an insightful view from both experimental and computational angle on a particular group of RR.


This comprehensive review describes in detail the well-studied fungal two-component system and its interaction with the MAPK-mediated HOG stress response signaling pathway.


41. Puthiyaveetil S, Kavanagh TA, Cain P, Sullivan JA, Newell CA, Gray JC, Robinson C, van der Giezen M, Rogers MB, Allen JF: **The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts.** Proc Natl Acad Sci USA 2008, 105:10061-10066. This work describes identification of a nuclear encoded HK in Arabidopsis whose protein product is exported to the chloroplast. The protein is shown to be evolutionarily related to plastid encoded HKs in red algae and to cyanobacterial HKs.

Evolution and phyletic distribution of two-component signal transduction systems

Wuichet, Cantwell and Zhulin


