Science Signaling
AAAS

Origins and Diversification of a Complex Signal Transduction System in Prokaryotes

Kristin Wuichet and Igor B. Zhulin (29 June 2010)
Science Signaling **3** (128), ra50. [DOI: 10.1126/scisignal.2000724]

The following resources related to this article are available online at http://stke.sciencemag.org.
This information is current as of 6 July 2010.

Article Tools Visit the online version of this article to access the personalization and article tools:
http://stke.sciencemag.org/cgi/content/full/sigtrans;3/128/ra50

Supplemental Materials *"Supplementary Materials"*
http://stke.sciencemag.org/cgi/content/full/sigtrans;3/128/ra50/DC1

Related Content The editors suggest related resources on *Science*'s sites:
http://stke.sciencemag.org/cgi/content/abstract/sigtrans;3/128/eg5
http://stke.sciencemag.org/cgi/content/abstract/sigtrans;3/128/pe23

References This article has been **cited by** 1 article(s) hosted by HighWire Press; see:
http://stke.sciencemag.org/cgi/content/full/sigtrans;3/128/ra50#BIBL

This article cites 85 articles, 45 of which can be accessed for free:
http://stke.sciencemag.org/cgi/content/full/sigtrans;3/128/ra50#otherarticles

Glossary Look up definitions for abbreviations and terms found in this article:
http://stke.sciencemag.org/glossary/

Permissions Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl

Downloaded from stke.sciencemag.org on July 6, 2010

Science Signaling (ISSN 1937-9145) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Copyright 2008 by the American Association for the Advancement of Science; all rights reserved.

EVOLUTION

Origins and Diversification of a Complex Signal Transduction System in Prokaryotes

Kristin Wuichet[1,2] and Igor B. Zhulin[1,2]*

(Published 29 June 2010; Volume 3 Issue 128 ra50)

The molecular machinery that controls chemotaxis in bacteria is substantially more complex than any other signal transduction system in prokaryotes, and its origins and variability among living species are unknown. We found that this multiprotein "chemotaxis system" is present in most prokaryotic species and evolved from simpler two-component regulatory systems that control prokaryotic transcription. We discovered, through genomic analysis, signaling systems intermediate between two-component systems and chemotaxis systems. Evolutionary genomics established central and auxiliary components of the chemotaxis system. While tracing its evolutionary history, we also developed a classification scheme that revealed more than a dozen distinct classes of chemotaxis systems, enabling future predictive modeling of chemotactic behavior in unstudied species.

INTRODUCTION

Three major modes of signal transduction in prokaryotes are recognized on the basis of the design of the regulatory system. The simplest signal transduction systems consist of a single protein, which is capable of both sensing a signal and directly affecting a cellular response, for example, a ligand-binding transcriptional regulator. Such proteins, termed one-component systems (*1*), typically use two separate domains: input (also called a sensory domain) and output (also called a regulatory domain). A more complex mode of prokaryotic signal transduction involves two functionally dedicated proteins, a sensor and a response regulator, that make up a two-component system (*2*). The sensor is a histidine kinase, which consists of an input domain and a transmitter domain that communicates (by means of phosphorylation) with the receiver domain of the response regulator, which in turn activates the response regulator's output domain. One- and two-component systems share a repertoire of input and output domains, but the main difference is that most one-component systems are known or predicted to detect signals in the cytoplasm, whereas most two-component systems are known or predicted to detect extracellular signals (*1*, *2*). Both one- and two-component systems primarily regulate gene expression through their DNA binding output domains (*1*, *2*), but they can also control other cellular activities through different types of output domains, such as cyclases, phosphodiesterases, and phosphatases (*1*, *3*). Variation in component design can be seen in both one- and two-component systems. For example, in one-component systems, a single-domain protein can be a sensor and a regulator (*4*), or multiple sensory and regulatory domains can be present in a single protein (*5*). In two-component systems, multiple sensory and regulatory domains per system can also exist (*3*, *6*), and additional phospho-acceptor and phospho-donor proteins can extend the system into a more complex phospho-relay (*7*).

The chemotaxis system, which is a special case of two-component signal transduction, constitutes the third mode. Bacteria navigate in chemical gradients by regulating their flagellar motility (*8*). This behavior, known as chemotaxis, is characterized by high sensitivity and precise adaptation, properties attributed to an assortment of interactions within the multi-protein signal transduction system (*8*, *9*). Although using principal components typical of two-component systems, its design is markedly different. The chemotaxis signal transduction system is best understood in *Escherichia coli* (Fig. 1). The histidine kinase of this system, the CheA protein, is sensorless (no input domain), and the cognate response regulator, the CheY protein, lacks an output domain (*10*, *11*). Although the sequence similarity of the CheA-CheY pair to classical two-component systems was noted (*11*), the CheA structure revealed such marked deviation from other known histidine kinases that CheA was proposed to constitute a separate class of histidine kinases, class II (*12*). All other histidine kinases were assigned to class I. CheA receives signals from dedicated chemoreceptors [also called methyl-accepting chemotaxis proteins (MCPs)] that are connected to the kinase through a docking protein, CheW, thus forming a signaling complex (*8*).

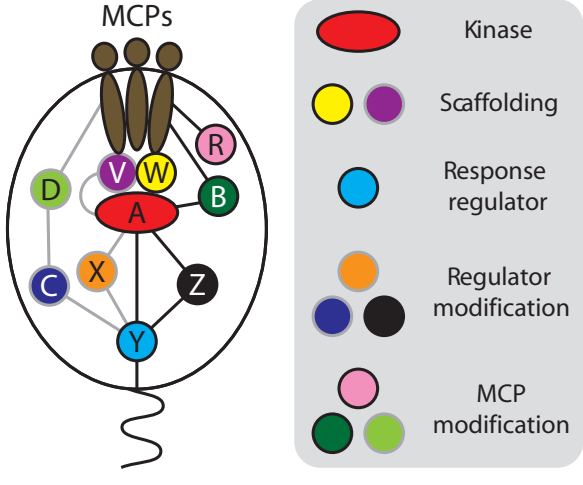


Fig. 1. Summary of current knowledge of organization of the bacterial chemotaxis system. Components outlined in light gray and their interactions (gray lines) are not present in the model organism *E. coli*, but have been studied in other species (*8*, *13*). A, CheA histidine kinase; W, CheW, and V, CheV scaffolding proteins; Y, CheY response regulator; R, CheR methyltransferase; B, CheB methylesterase; Z, CheZ phosphatase; C, CheC phosphatase; X, CheX phosphatase; D, CheD deamidase.

[1]BioEnergy Science Center and Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [2]Department of Microbiology, University of Tennessee, Knoxville, TN 37996, USA.
*To whom correspondence should be addressed. E-mail: ijouline@utk.edu

Downloaded from stke.sciencemag.org on July 6, 2010

MCPs, CheW, CheA, and CheY encompass an excitation pathway in chemotaxis. A dedicated phosphatase, CheZ (which acts on CheY), and the CheR methyltransferase and CheB methylesterase (which covalently modify MCPs) constitute an adaptation pathway. Thus, the model chemotaxis system includes seven different types of proteins, making it the most complex signal transduction system in prokaryotes with respect to component design.

Similar systems were identified in dozens of species other than *E. coli* (*8*, *13*). Most experimentally studied systems control flagellar motility (*13*), whereas some regulate motility based on type IV pili (Tfp) (*14*, *15*) and other cellular functions, such as development (*16*, *17*), biofilm formation (*18*), cell morphology, cell-cell interactions (*19*), and flagella biosynthesis (*20*). Although some of its components exhibit similarities to known regulatory proteins, many elements of the chemotaxis system have not been found in any other type of signal transduction system, and its overall design appears to be substantially more complex than that of any other signal transduction system in prokaryotes. Emerging models for studying the bacterial chemotaxis system, such as *Bacillus subtilis*, *Borrelia burgdorferi*, and *Helicobacter pylori*, lacked components found in *E. coli*—the CheZ phosphatase or the CheB and CheR methylation regulation enzymes—and revealed chemotaxis components that were not seen in *E. coli*—CheC and CheX phosphatases that act on CheY (*21*), the CheV docking protein, or the CheD deamidase that acts on MCPs, or some combination thereof (*14*, *22*). Other model organisms also revealed some variation in the number of protein components per system and their domain architecture, as well as the presence of multiple chemotaxis systems in a single organism. For example, *Rhodobacter sphaeroides*, *Myxococcus xanthus*, and *Pseudomonas aeruginosa* all have multiple sets of chemotaxis genes that constitute multiple chemotaxis systems with defined functions (*15*, *23*, *24*). However, compared to the millions of microbial species inhabiting our planet, these are few examples. The degree of diversification of the chemotaxis system remains largely unexplored and its origins are unknown. We took advantage of the wealth of genomic data to further explore the diversity of the chemotaxis system and gain insight into its origins.

RESULTS

Genomic view of the chemotaxis system

Using homologs of the known chemotaxis proteins as queries, we searched a nonredundant set of 450 prokaryotic genomes (Materials and Methods) and identified chemotaxis proteins in genomes of 245 species representing 16 phyla of Bacteria and Archaea, but not Eucarya (table S1 and Fig. 2). Database searches with CheB and CheR revealed a subset of protein sequences that had some features of CheB and CheR but were not predicted to be part of chemotaxis systems (table S2). These included sequences that were (i) lacking a domain that is part of CheB or CheR (such as a receiver domain in CheB), (ii) integral parts of larger proteins with extra domains not typical of CheB and CheR (such as a histidine kinase phosphotransfer domain), and (iii) found in genomes without any other chemotaxis genes. Subsequent phylogenetic analysis showed that these sequences formed a highly conserved clade (fig. S1), and these proteins were not considered members of chemotaxis systems in subsequent analyses. Instead, they were examined separately, which led to the identification of previously unknown classes of signal transduction proteins.

To address larger evolutionary questions about the chemotaxis system, we first addressed the basic question: Which elements of the chemotaxis system are present in all chemotactic organisms and which can be lost or substituted with other components during evolution? By measuring the frequency of chemotaxis gene occurrence (Fig. 3A) and co-occurrence (Fig. 3B), we established that MCPs, CheA, and CheW are present in >95% of genomes that contain at least one chemotaxis gene (Fig. 3A). Note that CheY is not included because it cannot be distinguished from other single-domain response regulators on the basis of sequence alone. Of the 245 genomes that encode any type of protein predicted to be part of a chemotaxis system, only seven lacked MCPs, CheA, or CheW, or some combination thereof. Three of these genomes (*Bacillus thuringiensis*, *Bacillus weihenstephanensis*, and *Listeria monocytogenes*) lack CheW but have the CheV protein, which is a fusion of CheW with a response regulator receiver domain, suggesting a chemotaxis system in which CheV substitutes for CheW. Experimental studies in *L. monocytogenes* confirm that its chemotaxis system is functional (*25*). Four other genomes (*Campylobacter concisus*, *Hyphomonas neptunium*, *Novosphingobium aromaticivorans*, and *Sodalis glossinidius*) lack MCPs, CheA, or CheW, or some combination thereof (table S1), and they are not predicted to form a functional chemotaxis system. Thus, MCPs, CheA, and CheW are common to chemotactic prokaryotes and represent part of the central chemotaxis system core. This evolutionarily irreplaceable set is also likely to include an output protein (CheY), thus representing a complete excitation pathway for chemotaxis (*8*).

We further inferred that CheB- and CheR-mediated demethylation and methylation are the dominant form of chemotactic adaptation (Fig. 3). CheB and CheR are found in ~90% of genomes with chemotaxis components (Fig. 3A), and the co-occurrence analysis shows that 90% of genomes that contain MCPs, CheA, and CheW also encode CheB and CheR proteins (Fig. 3B). Thus, we designate CheB and CheR (along with MCPs, CheA, CheW, and CheY) as central components of the chemotaxis system. The fact that CheB and CheR are missing from genomes of several chemotactic species (*14*, *26*) is consistent with experimental evidence that they are essential for some types of chemotaxis, but dispensable for others. For example, in *E. coli*, a *cheB/cheR* double mutant is null for chemotaxis to amino acids, but exhibits nearly normal chemotaxis to oxygen and redox signals (*27*).

The auxiliary proteins (CheC, CheD, CheV, CheX, and CheZ) are found in <60% of these genomes (Fig. 3). CheC, CheX, and CheZ proteins are all CheY phosphatases; however, even if pooled together in genomic counts, >20% of genomes with chemotaxis components do not have any identifiable CheY phosphatase. This suggests the evolutionary variability and perhaps even dispensability of phosphatases and further justifies the inclusion of CheC, CheX, and CheZ into the auxiliary component set. Consistent with their roles as auxiliary components, we found CheC, CheD, CheV, CheX, and CheZ exclusively in species that contain at least one central component.

We used the genomic analysis to address universal questions about the chemotaxis system. How widespread is chemotaxis in the microbial world? We show that more than half of sequenced prokaryotic genomes contain varying sets of chemotaxis genes (Fig. 4). This number slowly grows with the number of sequenced genomes. These data, along with the phyletic distribution of chemotaxis systems (Fig. 2), suggest that most prokaryotic species are chemotactic. Are multiple chemotaxis systems in a single organism, such as seen in *R. sphaeroides* and a few other species (*15*, *24*), an exception or the rule in the microbial world? We found that multiple chemotaxis systems occur as frequently as single ones (Fig. 5), which highlights the importance of investigating microbial models with multiple chemotaxis systems.

Identification of coevolving components and construction of the chemotaxis system tree

We developed a systems-level phylogenomic approach to the classification of the chemotaxis system to reveal its diversity. This approach contains

Downloaded from stke.sciencemag.org on July 6, 2010

three principal stages: (i) identification of coevolving central chemotaxis components and construction of their phylogenetic tree, (ii) matching nodes and leaves on the tree to distinct phylogenomic markers to define classes of the chemotaxis system, and (iii) assigning auxiliary components to the classes. As the first step in the classification process, we used classical phylogenetic techniques. The phylogenetic signal can be substantially increased by building concatenated multiple sequence alignments of proteins participating in the same function. This process was effective in a large-scale analysis of the tree of life (*28*), as well as in studies focused on a single cellular function (*29*). The central chemotaxis components present in all genomes were the obvious choice for building such an alignment; however, MCPs typically are present in more than one copy per system (*30*), and more than one copy of CheW per system is also common (table S1). Multiple copies of these proteins suggest functional diversification within a single system, and this disqualifies MCPs, CheW, and CheY from inclusion in the concatenated alignment. Therefore, only CheA appeared to be a suitable phylogenetic marker. Although two *cheA* genes were identified in one of the chemotaxis gene clusters in *R. sphaeroides* (*24*), which raises a question regarding the suitability of CheA as a phylogenetic marker of an individual chemotaxis system, neither of the two *cheA* sequences encodes a fully functional CheA protein, and both are required for a proper function of the single system. We have identified 11 instances of split CheA sequences (compared to 470 complete CheA sequences) in our genome set (table S1) and only used the portion encoding the globular CheA core, identified in its crystal structure (*12*), in phylogenetic analysis (Materials and Methods).

We included CheB and CheR in a concatenated alignment, which increased the phylogenetic signal at the expense of not including systems lacking CheB and CheR that constitute <10% of all systems in the genomic data set. Cognate CheA, CheB, and CheR sequences were identified by coevolution analysis (see fig. S2 for methods flow chart). We collected sequences belonging to the same system from genomes containing only one copy each of CheA, CheB, and CheR. In genomes containing multiple chemotaxis systems, coevolving (and therefore belonging to the same system) sequences were identified by gene neighborhood analysis: The presence of functionally related genes in the same gene neighborhood strongly suggests their involvement in the same process and coevolution (*31*). We found that >50% of coevolving CheA, CheB, and CheR sequences were in the same gene neighborhoods (table S1 and fig. S2), and >70% of CheA sequences had a coevolving CheB or CheR, or both, encoded in their gene neighborhoods. The rest of cognate CheA, CheB, and CheR were assigned on the basis of sequence similarity with mirror tree analysis. Briefly, individual phylogenetic trees of CheA, CheB, and CheR (fig. S1) were examined for overall similarity and consistency of individual clades (for example, organism distributions within given clades, branching patterns within and between the clades, and branch lengths) that reflect coevolution of protein families (*32*, *33*). CheA, CheB, and CheR were found in a nearly 1:1:1 ratio within the genomic data set, making this analysis fairly straightforward: >80% of all sequences were assigned to coevolving sets (fig. S2).

Defining chemotaxis system classes with phylogenomic markers

We constructed a maximum likelihood tree from the concatenated multiple sequence alignments of the cognate CheA, CheB, and CheR sequences (Materials and Methods). The tree was rooted at its midpoint and then progressively analyzed node by node from its root to its leaves (Fig. 6 and figs. S3 and S4). At each internal node, we gathered the descendant sequences and matched them to unique markers to determine whether the descendants should be grouped into a single class or split into multiple classes. The phylogenomic markers included (i) gene order of the corresponding gene neighborhoods; (ii) auxiliary component content of the corresponding

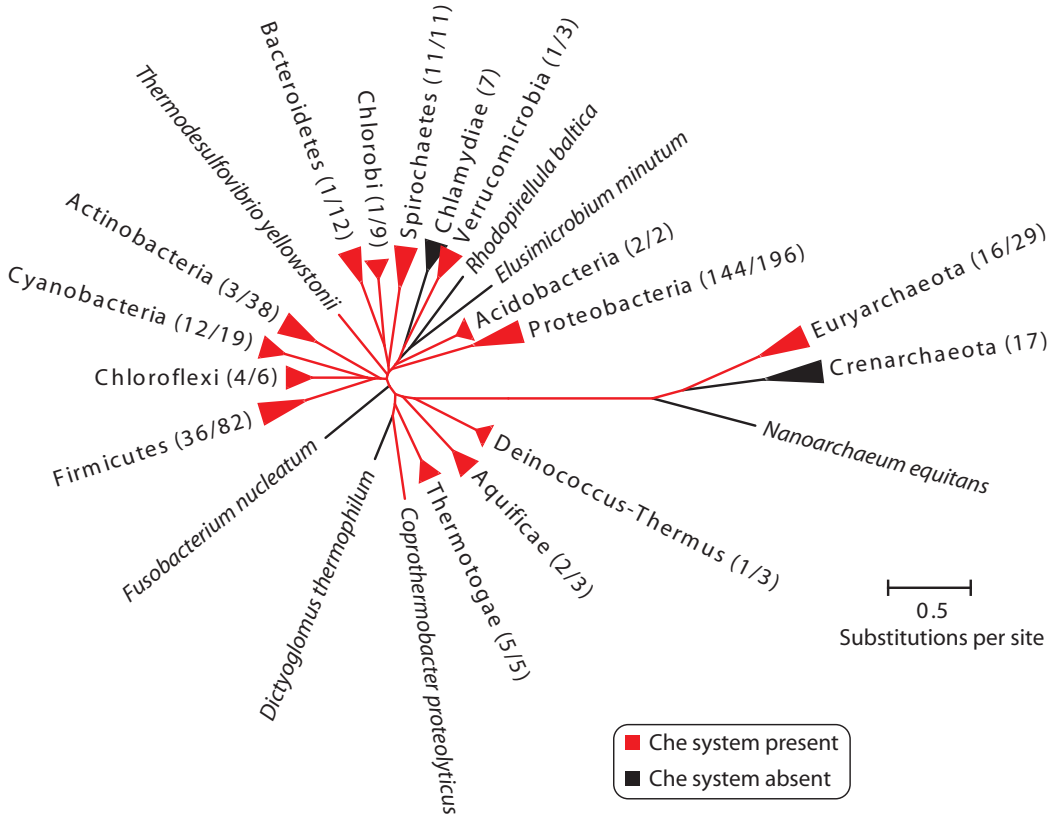


Fig. 2. Phyletic distribution of the chemotaxis system. Relationships between major prokaryotic phyla are shown as an unrooted maximum likelihood tree (Materials and Methods). Phyla containing representatives with chemotaxis system components are shown in red. The number of these representatives versus the total number of analyzed genomes within the clade is shown in parentheses. For phyla lacking chemotaxis (shown in black), only the number of genomes in the clade is shown.

Downloaded from stke.sciencemag.org on July 6, 2010

gene neighborhoods; (iii) domain architectures of CheA, CheB, CheR, and their gene neighboring chemotaxis components; and (iv) the signaling domain class of MCPs encoded in gene neighborhoods with CheA, CheB, or CheR (*30*). Gene order in prokaryotes is not conserved; therefore, patterns of gene order conservation are strong indicators of relatedness that are independent from sequence similarity (*34*, *35*). Chemotaxis components display various domain architectures (*10*), and we used noticeable deviations from standard architectures as markers to assign sequences that shared these deviations to a particular class (Fig. 6). To introduce yet another marker into our classification scheme, we took advantage of a previous study that established conserved classes of MCPs defined by sequence similarity and the number of helical heptads (for example, 40H class typically has 40 heptads) in their signaling domains (*30*). The progressive root-to-leaves analysis of these markers on the tree supported the existence of at least 18 classes within the chemotaxis system (Fig. 6).

In the next step of the classification process (fig. S5), sequences that were assigned to classes on the basis of the CheA-CheB-CheR tree and phylogenomic markers at the previous stage were marked on the individual phylogenetic trees of CheA, CheB, and CheR. This enabled us to classify the rest of CheA, CheB, and CheR sequences (that were not included in the concatenated alignment) on the basis of their similarity to the classified homologs. These newly classified sequences may belong to degenerate systems, which do not form a functional chemotaxis system, or functional systems that have undergone lineage-specific gene loss. For example, of the three well-studied *R. sphaeroides* chemotaxis systems, one lacks a cognate CheB (*24*), but its CheA and CheR components were assigned to a specific class on the basis of similarity to the classified sequences. Studies of the incomplete system in *R. sphaeroides* suggest it is degenerate (*24*), unlike the functional chemotaxis systems in *Synechocystis* sp. PCC 6803 that lack CheB and CheR (*14*, *36*). The CheA tree revealed one additional class of chemotaxis system lacking CheB and CheR, which was not one of the previously identified 18 classes. This concluded assignment of 98% of all available CheA, CheB, and CheR sequences to 19 distinct classes within the chemotaxis system (Fig. 7).

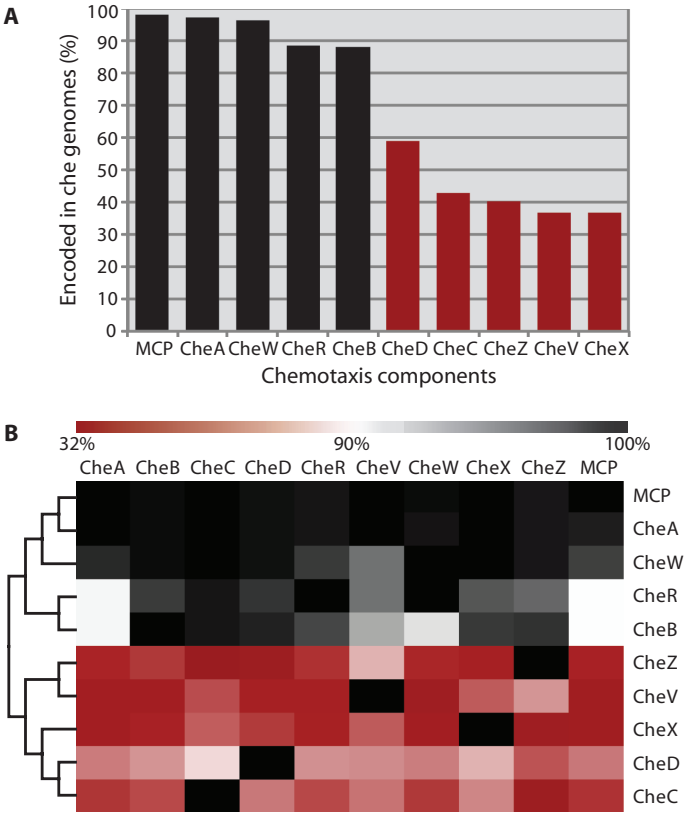


Fig. 3. Defining a minimum set, core, and auxiliary chemotaxis components. All genomes that have at least one chemotaxis component were taken into consideration. CheY proteins were not considered because of problems with their identification (Materials and Methods). (**A**) Relative frequency distribution of chemotaxis components in genomes. (**B**) Co-occurrence of chemotaxis components in genomes. Blocks show the percentage of genomes encoding the column component that also encode the row component according to the gradient colors at the top. The tree at the left shows the results of average linkage hierarchical clustering using Euclidean distance.

Assigning CheW and auxiliary components to classes

We assigned the CheC, CheD, CheV, CheW, CheX, and CheZ (CDVWXZ) sequences to classes by using genomic distribution, gene neighborhood, and sequence similarity analyses (fig. S6). Nearly 30% of all CDVWXZ sequences were found in genomes with a single CheA protein, which was classified in the previous steps, and therefore were automatically assigned to the same class. For example, the *H. pylori* genome, which encodes a single CheA protein, also has CheZ, which is found outside the chemotaxis gene neighborhood (*37*). Similarly, genes encoding the CheC, CheD, and CheX components of the single *Thermotoga maritima* chemotaxis system are also found outside the main chemotaxis gene cluster (*21*). More than 60% of the remaining sequences were found in gene neighborhoods that contained classified CheA, CheB, or CheR sequences, or some combination thereof. All CDVWXZ sequences that were encoded near CheA, CheB, or CheR sequences of a single class were automatically assigned to that same class. For example, the single CheZ protein of *P. aeruginosa* is encoded in one of the four chemotaxis gene neighborhoods in this organism and thus was assigned to the same class as the neighboring CheA protein (*38*). Finally,

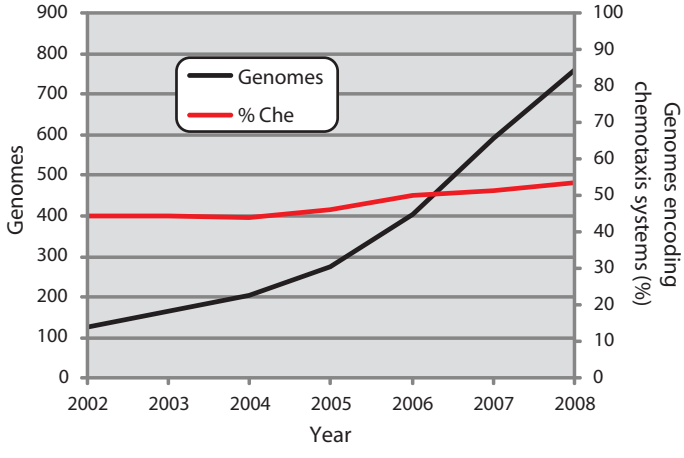


Fig. 4. Prevalence of chemotaxis systems in sequenced genomes. The graph includes complete genomes from all sequenced prokaryotic species available before November 2008, but the trend (a slight increase in the percentage of genomes with chemotaxis system with the increase in the number of sequences genomes) remains even when our nonredundant genome set (Materials and Methods) was used.

Downloaded from stke.sciencemag.org on July 6, 2010

the sequences assigned to classes were marked on individual CheC, CheD, CheV, CheW, CheX, and CheZ trees (fig. S7) to evaluate classifications and assign the remaining sequences to classes on the basis of sequence similarity to assigned homologs. Very few (<4% of all CDVWXZ sequences) that were assigned to a class in previous steps were reclassified because of relevant discrepancies (for example, sequence dissimilarity, taxonomic differences, or the lack of a cognate CheA, CheB, or CheR of the same class in the same organism).

We found that CheC and CheX sequences had sporadic genome distributions and inconsistent tree topologies; thus, almost half of these sequences remained unclassified. We chose to err on the side of caution when assigning these phosphatases to chemotaxis systems because they could also act on receiver domains of components that are not involved in chemotaxis. Indeed, we found a previously unidentified class of nonchemotaxis two-component system that has members predicted to use CheX fused to their kinases or encoded in their gene neighborhoods (table S3). The histidine kinases of these two-component systems have an intermediate domain architecture when compared to class I and class II (CheA) histidine kinases that we termed class III histidine kinases (HKIIIs), and their two-component systems are predicted to lack MCPs and CheW, which are essential chemotaxis components. Overall, ~90% of CDVWXZ components were classified into chemotaxis systems, which resulted in the classification of 93% of all chemotaxis components in the analyzed genomic data set. (The 90% does not include the CheX proteins assigned to nonchemotaxis systems, although they make up such a small fraction that the number would be virtually identical even if they were included.)

Determining relationships between chemotaxis system classes and behaviors

After completion of sequence-based classification, we collected available experimental evidence implicating individual chemotaxis proteins and systems in governing specific cellular functions (table S1) and mapped this information onto the individual CheA, CheB, and CheR trees and the tree built from the concatenated CheA-CheB-CheR alignment to assign potential functions to individual classes. The results supported the existence of three functional groups of chemotaxis system: those that regulate flagellar motility (Fla), those involved in Tfp-based motility (Tfp), and those with alternative (nonmotility) cellular functions (ACF). Response regulators of the ACF class were predominantly found to contain specific output domains that link these systems to alternative targets (Fig. 7 and table S1). Of the 19 identified classes, 17 belong to the Fla group (F1 to F17), whereas the Tfp and ACF groups each contain a single class (Fig. 7).

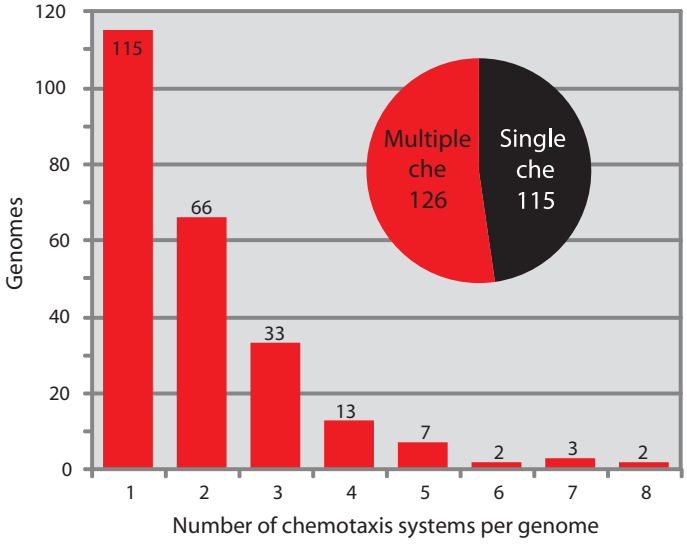


Fig. 5. Distribution of single and multiple chemotaxis systems. The number of chemotaxis (che) systems in our genome set is defined here as the number of CheA proteins it is predicted to encode. In instances of split CheAs, only one protein was counted.

Previously unknown intermediate forms between classical two-component systems and chemotaxis: MAC and HKIII

We analyzed the set of atypical CheB and CheR proteins that did not conform to canonical domain architecture and that formed distinct clades on respective phylogenetic trees (fig. S1). These CheB-CheR–like pairs were associated with several types of proteins that all contained predicted two-helix coiled-coil regions with recognizable methylation sites (Fig. 8 and figs. S8 and S9), which are conserved glutamate pairs typical of MCPs (*30*). We termed these proteins methyl-accepting coiled-coil (MAC) proteins. There were two types of MAC-CheB-CheR associations: Either CheB and CheR were integral domains of MAC proteins (termed MAC1 systems) or they were encoded in the same gene neighborhood with a MAC protein (MAC2 systems). MAC proteins were found in ~20% of all analyzed genomes (table S2), and 83% of genomes that encoded MAC proteins also contained chemotaxis components. The MAC module was found in proteins that control diverse functions including cyclic diguanosine monophosphate regulation and DNA binding; however, class I histidine kinases (HKIs) make up the output of 60% of MAC1 proteins and are the exclusive output of MAC2 proteins (table S2).

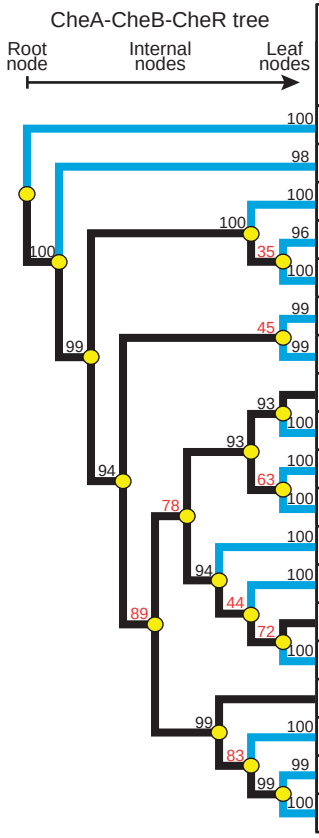
Upon the discovery of MACs, we also searched for other signal transduction proteins that were not predicted to be part of the chemotaxis system but might have some of its features. Using similarity searches initiated with CheA protein sequences, we found sensor histidine kinases with domain architectures that have features of both class I and class II (CheA), which we termed HKIIIs. HKIIIs have an N-terminal sensory module typical of class I, but they are predicted to autophosphorylate a histidine phosphotransfer (HPT) domain that is N-terminal to the kinase module, similarly to CheA (Fig. 9A), rather than at the dimerization domain as seen in HKIs. A multiple sequence alignment of HKIII shows that the only conserved histidine is in the HPT domain and none is found in the region corresponding to the dimerization domain (fig. S10). Sequence analysis also revealed that the N-box region of the kinase domain shows strong similarity to that of CheA and not to HKIs (Fig. 9B). Only 4% of analyzed genomes contained HKIIIs, and all of these genomes also contained chemotaxis genes. HKIIIs are often encoded near CheY-like proteins and CheX, a CheY phosphatase (table S3).

DISCUSSION

Independent phylogenetic (*39*) and genomic (*1*) studies established that two-component signal transduction originated in Bacteria and then was laterally transferred to Archaea. We suggest the same pattern for the chemotaxis system. We also propose that the chemotaxis system originated from classical two-component systems (*2*) through simple, incremental innovations, such as domain acquisition, protein recruitment, and a single instance of a new domain birth. The newly discovered class III and MAC systems that have features of both two-component and chemotaxis systems represent functionally intermediate forms, suggesting a gradual progression from classical two-component systems to the chemotaxis system. In addi-

Downloaded from stke.sciencemag.org on July 6, 2010

tion to the increased design complexity of the chemotaxis system compared to classical two-component systems, two other lines of evidence support this notion. First, two-component systems that are not associated with chemotaxis (defined by HKIs) are more widely distributed among prokaryotes than are the chemotaxis system: >95% of prokaryotic genomes contain two-component systems (*40*), whereas only 50% contain the chemotaxis system. Second, the repertoire of associated input domains is substantially more diverse in nonchemotaxis systems: Among the 30 most abundant input domains, 17 are not associated with chemotaxis and none is unique to chemotaxis systems (table S4).



Chemotaxis system class	Phylogenomic markers			
	Gene order conservation	GN auxilliary component conservation	Standard domain architecture deviations	MCP class
Tfp	YYWMRABW	none	HPT[5-6]:CheA-REC	40H
ACF	WRW...ABO	none	CheA:REC, CheR:TPR	40H
F7-major	YAW...R...B	D	None	36H
F7-divergent	None	none	None	36H
F8	None	D	None	34H
F1	None	C, D	None	44H
F2	AW:RXY RB	X	CheW:CheR	
F16	OYYAWMR...RB	none	None	Uncat
F11	WAMRB	none	MCP (long, no TM)	Uncat
F13	YAWM...RB	none	MCP (long, no TM)	Uncat
F10	YA...W...W...RB	none	None	
F9	MWB...RYA	none	MCP_signal:MCP_signal CheW:CheW:CheW	44H
F4	B...R...WYA	none	HPT[2]:CheA	
F17	RYYAW:ZB	W:Z	CheW:CheZ	
F12	WAR...B	none	None	
F15	WRYZAB...M	Z	None	40H
F3	VAW BR	V	CheA:REC, CheB (no REC)	
F5	AW...BR	none	CheA:CheW	38H
F6	YZAB...WW VR	Z, V	None	

Standard domain architectures

CheA - HPT…H-kinase_dim:HATPase_c:CheW	CheW - CheW
CheB - Response_reg:CheB_methylest	CheZ - CheZ
CheR - CheR_N:CheR	MCP - TM…TM…MCP_signal

Fig. 6. Defining chemotaxis system classes with phylogenomic markers. The CheA-CheB-CheR tree was rooted at its midpoint and then progressively analyzed node by node from its root nodes to its leaf nodes. At each internal node, sequences of each leaf were matched to markers to determine whether the descendants should be grouped into a single class (blue line) or split into multiple classes (yellow circle). aIRT scores from PhyML are shown at each internal node. An aIRT score of ≥90 is consistent with a bootstrap score of at least 75 in good-quality data sets (*85*), and poorly supported nodes with scores of <90 are shown in red. The markers shown in the table at the right include gene order, conserved auxiliary components in gene neighborhoods (GN), deviations from standard domain architectures of chemotaxis components in the gene neighborhoods, and the signaling domain class of neighboring MCPs (*30*). Gene orders and neighborhood components show CheA (A), CheB (B), CheC (C), CheD (D), CheR (R), CheV (V), CheW (W), CheX (X), CheY (Y), CheZ (Z), MCP (M), and non-CheY response regulators (O). The ":" symbol represents a fusion between components, and "…" represents one or more nonconserved or components not involved in chemotaxis, or both were found in between conserved che components or domains. The domains shown at the bottom correspond to Pfam domain models except for transmembrane regions (TM). In the table, the "Response_reg" model from Pfam has been shortened to "REC." A divergent subfamily of F7 systems (F7-divergent) groups with F8 systems in the CheA-CheB-CheR tree; however, the individual CheA tree shows all F7 systems grouped together (fig. S1), which is supported by the presence of a single MCP class (36H) in both the F7-major and F7-divergent classes.

The MAC module (CheB, CheR, methylated coiled coil) in histidine kinases represents a distinct step in increasing design complexity within two-component systems that has not been previously recognized. It may have evolved by extension of a sequence encoding generic α-helical region (such as a dimerization domain of histidine kinases) into a coiled coil and genetic recruitment of sequences encoding generic enzymes to produce modules with more than one catalytic domain and enzymatic activities in addition to the kinase activity. Indeed, the CheR methyltransferase is homologous to diverse DNA methylases (*41*). Progression from MAC systems to a core chemotaxis system can be achieved by only two events. First, the sensor and kinase domains of MAC split into an MCP and a sensorless kinase, and then a new domain, which is found as a domain in CheA kinases (Fig. 9) and also as an independent scaffolding protein, CheW, linking it to MCPs, is born. CheW belongs to a unique domain superfamily of the OB fold that is exclusively found in chemotaxis systems. The previous suggestion that CheW is reminiscent of a eukaryotic Src homology 2 domain (*12*) is not corroborated by current fold classifications, such as SCOP (Structural Classification of Proteins) (*42*). This suggests that the birth of the CheW domain has led to a transition from classical two-component systems that link multiple inputs (sensory domains of histidine kinases) to multiple outputs (output domains of cognate response regulators primarily targeting multiple DNA promoters) to the system that links multiple inputs (sensory domains of MCPs) to a single output (CheY to the flagellar motor). Thus, it is likely that the chemotaxis system has evolved as a specific benefit of controlling flagellar motility. It is important to stress that the proposed evolutionary scenario is hypothetical. Although MAC systems are widespread and found among diverse organisms (table S2), CheB and CheR sequences of these systems do not form taxonomically conserved clades in phylogenetic trees (fig. S1), which suggests that MAC systems are prone to horizontal transfer events and makes their evolutionary history difficult to

Downloaded from stke.sciencemag.org on July 6, 2010

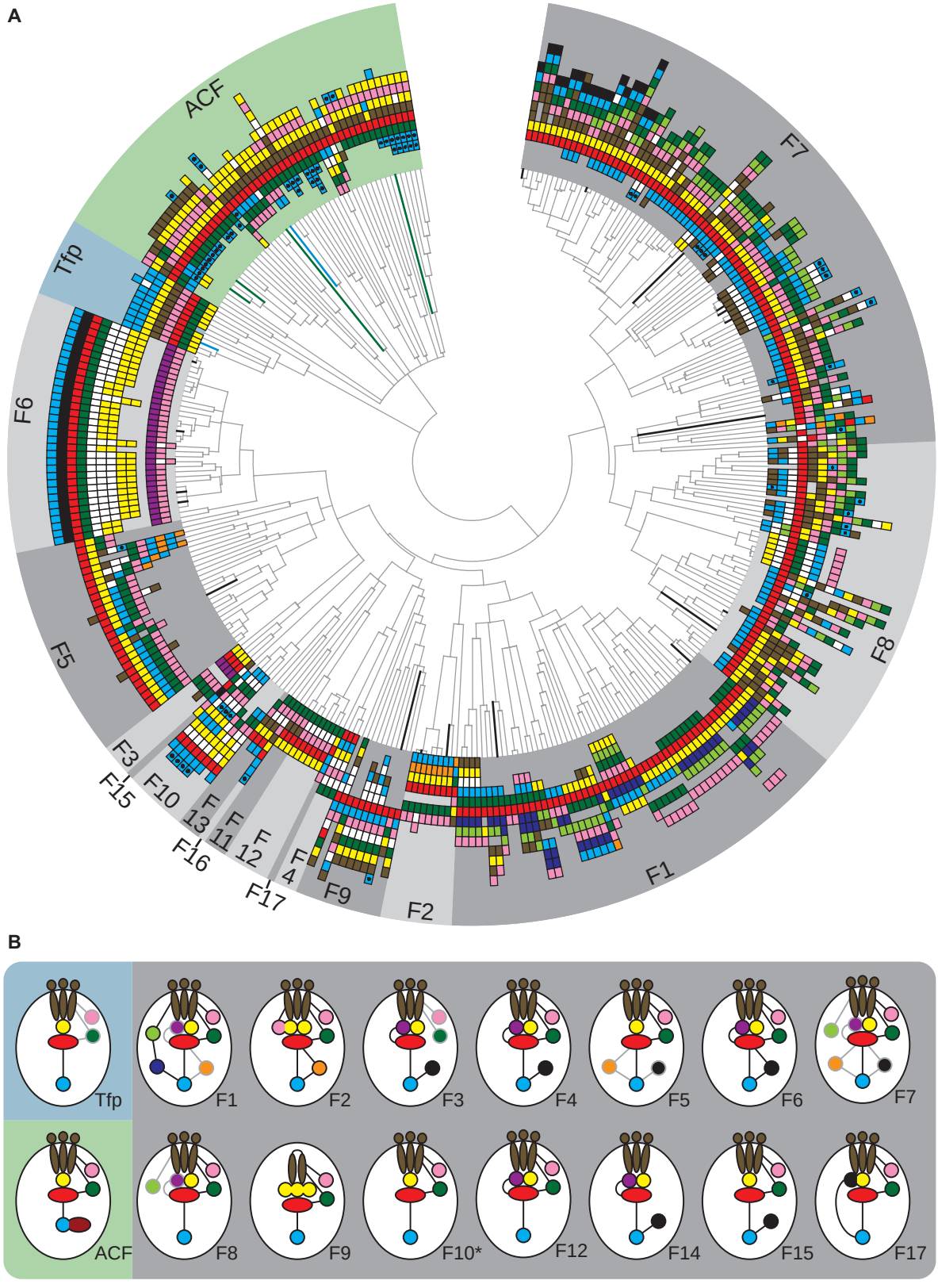


Fig. 7. Phylogenomic classification of the chemotaxis system. (**A**) A maximum likelihood phylogenetic tree built from concatenated multiple sequence alignment of CheA-CheB-CheR proteins is shown in the middle. Branches corresponding to proteins that have been experimentally shown to control flagellar motility are in black, Tfp motility are in blue, and alternative cellular functions are in green (table S1). The colorful wide concentric circle around the tree shows the gene neighborhood for genes corresponding to the CheA, CheB, and CheR protein sequences on the tree. Each gene is shown as a small colored rectangle. The color scheme is the same as in Fig. 1. Background color highlights 18 classes of the chemotaxis system: ACF, green; Tfp, blue; and 16 Fla classes, light and dark gray. The F14 class is not represented because it lacks CheB and CheR. (**B**) Protein interaction networks reconstructed for all 19 classes of the chemotaxis system. Color code is the same as in (A). F10* represents F10, F11, F13, and F16 interaction networks. Components with outlines and interaction lines shown in light gray are not common to all members of a class.

Downloaded from stke.sciencemag.org on July 6, 2010

explore. MAC systems could predate chemotaxis systems, consistent with their simpler component design; however, it is also plausible that CheB and CheR were recruited by MAC systems after the birth of the chemotaxis system. HKIIIs are an attractive putative evolutionary "missing link" for the transition from HKI to CheA because of their association with chemotaxis-like components, sequence similarities to CheAs, and their intermediate domain architecture and phosphorylation mechanism (table S3 and Fig. 8). However, the sparse distribution of HKIIIs prevents us from confidently describing them as anything beyond functional intermediaries.

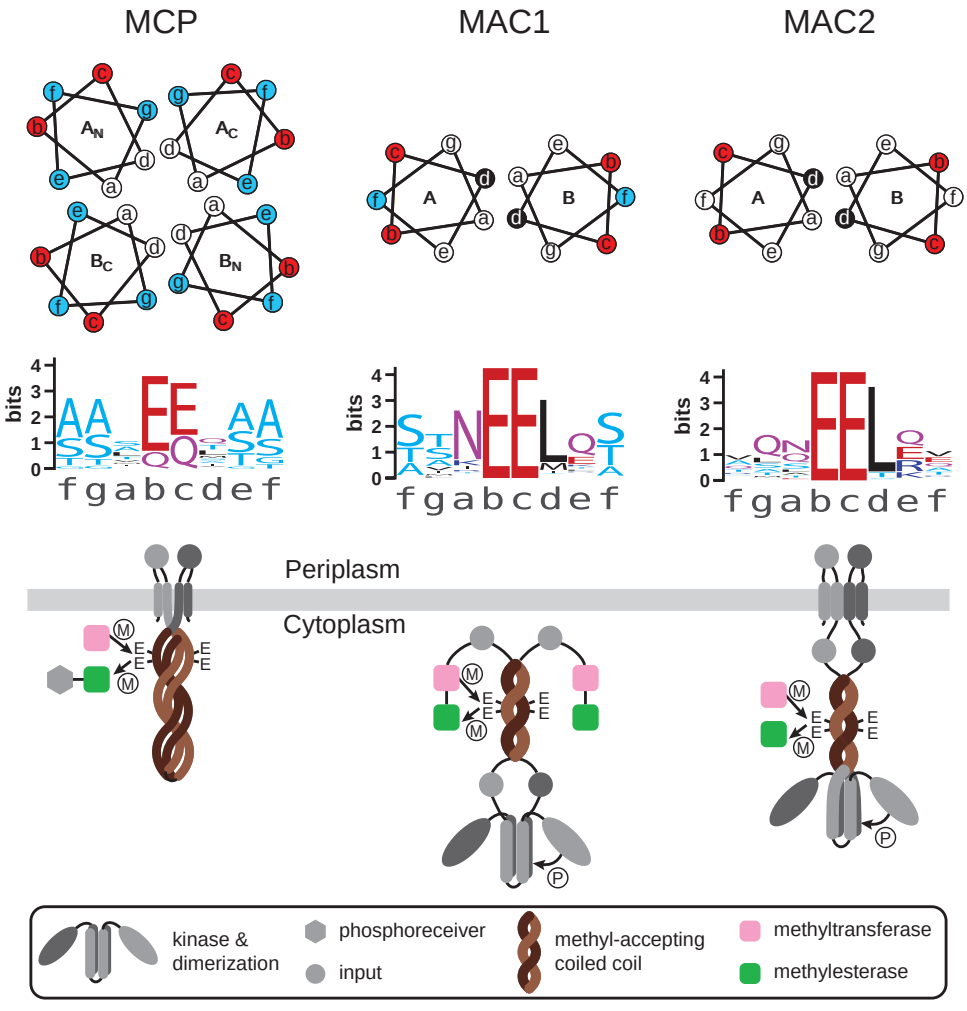


Fig. 8. Similarities between chemoreceptors and MAC kinases. At the top, helical wheel diagrams show the chemoreceptor methylation domain, which consists of a four-helix antiparallel bundle with two helices (N and C) from each monomer (A and B), and the MAC methylation region that is predicted to be a parallel two-helix coiled coil with a helix from each monomer (A and B). Helical wheel positions are colored according to conserved small (green), glutamate (red), and leucine (black) residues shown in the corresponding sequence logos below. Sequence logos (*86*) show the putative methylation consensus sequences for chemoreceptors, MAC1, and MAC2 systems. The α helix heptad positions, which match the helical wheels, are shown immediately below the logos. The chemoreceptor sequence logo was built from the signaling domains of the major class receptors identified in previous work (*30*), which match the established [AGST]-[AGST]-X-[EQ]-[EQ]-X-[AGST]-[AGST] consensus sequence. Logos for the MAC methylation sites were built from heptads of the methylation regions with glutamate (E) in the b and c positions (figs. S8 and S9). Schematic domain representations of chemoreceptor, MAC1, and MAC2 systems are shown at the bottom. The dimeric proteins are shown with light- and dark-colored monomers. For clarity, the methylation (M) and phosphorylation (P) interactions of only one monomer are shown for each dimer. Protein domains that are common to both classic two-component and chemotaxis systems are shown in gray, whereas other colors represent components exclusive to chemotaxis and MAC systems.

The three functional groups of the system (Fla, Tfp, and ACF) appear to have distinct evolutionary paths; however, because deep branches of the prokaryotic tree cannot be resolved (*43*, *44*), we cannot argue that one or another functional group is ancestral on the basis of their phyletic distribution. ACF systems might be an ancestral group because of their simple design and similarity of targets to those of the intermediate MAC systems. As the motility functions arose, these systems might have diversified and spread across the prokaryotic lineages. An equally plausible scenario places the major group Fla as evolutionarily the oldest because it shows the greatest diversity and the widest phyletic distribution of the three functional groups: Fla is found in 12 prokaryotic phyla, whereas Tfp and ACF are limited to 6 phyla. If the latter scenario is correct, the F1 class of the chemotaxis system [found almost exclusively in Firmicutes, Thermotogae, and Archaea that share many genes through horizontal gene transfer (*45*)] might be the most ancient. The fact that evolutionary genomics places chemotaxis systems of Firmicutes and Archaea in one class explains the previously observed paradox that the biochemistry and physiology of chemotaxis in the bacterium *B. subtilis* is similar to that of the archaeon *Halobacterium salinarum* and different from that of other bacteria, such as *E. coli* (*13*). Lineage-specific gene loss appears to play an important role in the diversification of the chemotaxis system. Cases of recent gene loss can be seen in proteobacteria. For example, *E. coli* lacks the CheV protein, which is present in other chemotactic enterobacteria (*Salmonella*, *Yersinia*, *Enterobacter*, *Erwinia*, and *Pectobacterium* species). Parsimonious interpretation of such a distribution is that *E. coli* has lost the gene rather than it independently evolved or was laterally transferred into all closely related species. Similarly, the experimentally studied *H. pylori* chemotaxis system (*26*) lacks CheB and CheR proteins, whereas most other chemotactic ε-proteobacteria, including the closely related *H. hepaticus*, have these proteins.

The central elements of the chemotaxis system that constitute the excitation (MCPs, CheA, CheW, and CheY) and adaptation (CheB and CheR) pathways in the model organism *E. coli* are conserved throughout the evolutionary history of prokaryotic chemo-

Downloaded from stke.sciencemag.org on July 6, 2010

taxis. The choice of *E. coli* as a model for chemotaxis (*46*) was unknowingly excellent: The presence of all central components and the lack of most auxiliary components subsequently enabled the detailed molecular mechanisms of chemotaxis to be determined. Although the chemotaxis systems in some organisms lack CheB and CheR, such systems usually belong to classes that typically contain these proteins. F1, F3, and Tfp systems that lack CheB and CheR are functional for chemotaxis on the basis of experimental studies of systems of the respective classes in *M. xanthus*, *H. pylori*, and *Synechocystis* sp. PCC6803 (*14*, *47*, *48*). It is commonly accepted that bacteria use temporal sensing rather than spatial sensing because of their small size (*49*), but the existence of functional chemotaxis systems lacking CheB and CheR suggests that these organisms may use spatial sensing or an alternative mechanism of adaptation. The much slower time scale of surface gliding in comparison to swimming confers the extra time needed for signal integration and spatial sensing in the case of Tfp-based cyanobacterial gliding motility (*50*, *51*). *H. pylori* colonizes the thick gastric mucosa, which may also confer a slower time scale compared to swimming in a less viscous environment and may thus allow for spatial sensing. However, the *H. pylori* chemotaxis system uses CheV proteins that confer a feedback loop analogous to the classical CheB-CheR adaptation pathway. CheV is a scaffolding protein homologous to CheW, and it also contains a response-regulator domain, similarly to CheB, which is phosphorylated by CheA. The *H. pylori* chemotaxis system also includes CheZ, which has been suggested to provide another means of adaptation (*52*). Furthermore, experimental studies in mutant strains of *E. coli* (*53*, *54*) and wild-type *R. sphaeroides* (*55*) have suggested that adaptation may not be essential for effective chemotaxis, a finding supported by in silico analysis (*56*). Despite these observations, the evolutionary conservation of CheB and CheR in most chemotaxis systems suggests that methylation-based adaptation is an effective fitness strategy for a variety of organisms and conditions.

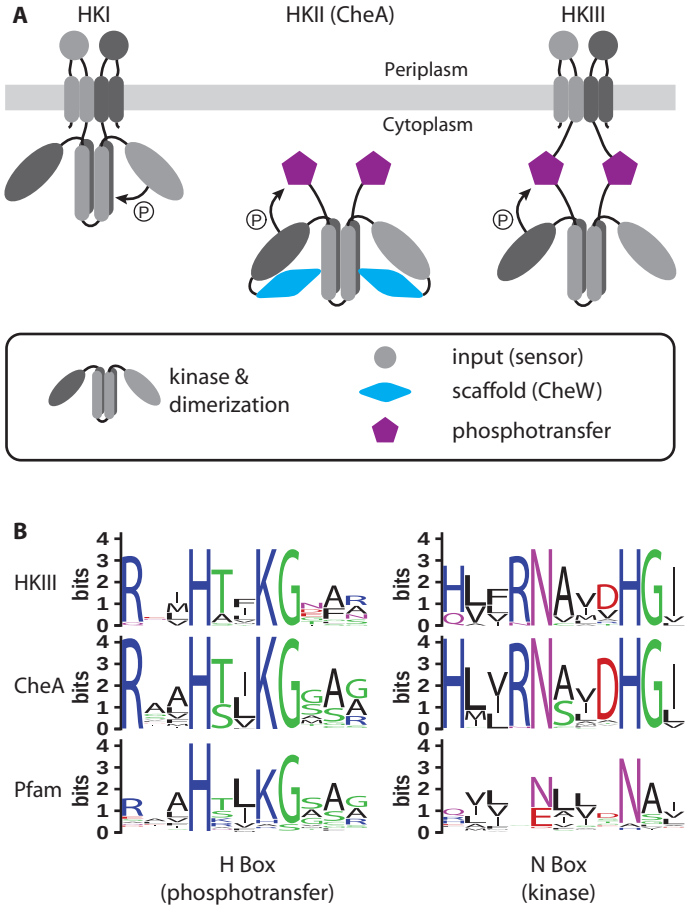


Fig. 9. HKIIIs have an architecture intermediate between those of class I and CheA histidine kinases. (**A**) Schematic domain representations of the three classes of kinases are shown at the top. The dimeric proteins are shown with light- and dark-colored monomers. For clarity, the phosphorylation interactions (P) of only one monomer are shown for each dimer. Protein domains that are common to both class I and class II (CheA) kinases are shown in gray, whereas other colors represent elements typically found in CheA and HKIII histidine kinases. (**B**) Although the HPT domain model (Pfam:Hpt) is not always identified in HKIII proteins, the sequence logos from the putative phosphorylation site (the only conserved histidine in HKIIIs; fig. S10) support the notion that the region is a phosphotransfer domain given the similarities between sequence logos (*86*) built from the CheA Hpt domain and the Pfam seed alignment used to build the Hpt domain model. The N box (*2*) of the HKIII ATPase domains also shows similarity to the N box of CheA kinases, more so than the N box of the Pfam seed alignment used to build the HATPase_c domain model. Although a variety of histidine kinases are represented by the HATPase_c domain model including CheA kinases and DNA gyrases, most of the seed alignment members are HKIs. The H- and N-box sequence logos from the HKIII sequences correspond to positions 54 to 64 and 274 to 284, respectively, of the HKIII alignment in fig. S10.

Special care needs to be taken when extrapolating chemotaxis system classification onto bacterial physiology. Groupings and classes presented here are a reflection of the evolutionary history of the chemotaxis system, and not necessarily the exact predictors of function. Whereas most members of any given class are likely to conform to the original function (for example, regulation of flagellar rotation), one should expect examples of neofunctionalization. This should be especially true for organisms with multiple chemotaxis systems that originated by gene duplication and horizontal gene transfer, the major driving forces in the evolution of new function in prokaryotes (*57*). For example, the F5 system in *Rhodospirillum centenum* regulates flagellar motility (*58*); however, its F9 system has a new function of controlling flagellar biosynthesis (*20*). Similarly, *M. xanthus*, which uses Tfp-based motility, lacks a Tfp chemotaxis system and has multiple, primarily paralogous ACF systems (table S1) that play roles in Tfp-based motility (*15*, *59*).

The classification work provides new targets for experimental study, because members of only two classes have been studied thoroughly (*E. coli* of the F7 class and *B. subtilis* of the F1 class) and members of nine classes have not yet been studied. Our analysis has just begun to uncover the true diversity of this function in prokaryotes. Upcoming genomic and especially metagenomic data sets will reveal more types and classes of the chemotaxis system. We do not know how differences between classes of the chemotaxis system affect the resulting chemotaxis behavior. Future experimental and theoretical studies will address this question.

The birth of the chemotaxis system represents a major transition in signal transduction from passive sensing to active exploration of the environment. None of the components of the bacterial chemotaxis system are present in eukaryotes, indicating that prokaryotes and eukaryotes developed navigation systems that are similar in basic principles (use of dedicated receptors and protein kinases) but markedly different in component design [for example, MCPs versus G protein (heterotrimeric guanine nucleotide–binding protein)–coupled receptors, histidine kinases versus mitogen-activated protein kinases]. However, the analogy between the bacterial chemotaxis system and the nervous system in eukaryotes that was

Downloaded from stke.sciencemag.org on July 6, 2010

previously drawn purely on the behavioral level (*60*) can now be extended. Both systems are pinnacles of an independent complex development of regulatory systems in the two major lineages of life.

MATERIALS AND METHODS

Bioinformatics software and computer programming environment

The following software packages were used in this study: Gblocks v0.91b (*61*), HMMER v2.3.2 (*62*), MAFFT v6.240 (*63*), MEGA v4.0 (*64*), Phylip v3.67 (*65*), PhyML v3.0 (*66*), PSI-BLAST (*67*) from BLAST v2.2.17, and VISSA (*68*). All multiple sequence alignments were built in MAFFT with its l-INS-i algorithm unless otherwise specified. All neighbor-joining trees were built in MEGA with pairwise deletions and the JTT substitution matrix. All maximum likelihood phylogenetic trees were built in PhyML with LG + Γ_4 + F parameters and subtree pruning and regrafting topology search with approximate likelihood ratio tests (alRTs) to evaluate branch support unless otherwise specified.

All computational analyses were performed in a local computing environment (including high-performance computing), and custom scripts for data analysis were written in Perl. A local version of the MiST (Microbial Signal Transduction) database (*40*) was available for direct queries using Perl scripts. The MiST database implements all Pfam domain models, but it specifically focuses on cataloging the three major prokaryotic signal transductions systems (one-component, two-component, and chemotaxis systems) on the basis of domain architecture queries with publicly available and privately curated domain models.

Data sources

Proteomes of each distinct prokaryotic species for which at least one complete genome sequence was available in National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) database (*69*) as of October 2008 were collected (503 proteomes). Redundant species in which the ribosomal RNA and ribosomal protein sequences were nearly identical [for example, *Bordetella pertussis* and *Bordetella parapertussis* (*70*)], but which have different species names because of phenotypic reasons, were identified and excluded. To do this, we built a multiple sequence alignment of the L5 ribosomal sequences in MAFFT and then used it to construct a distance matrix with the JTT amino acid substitution matrix and default parameters of Protdist in the Phylip package. Groups of sequences for which pairwise distances among all members were <0.023 were identified. These groups corresponded to clades within a neighbor-joining tree built from the distance matrix by means of Neighbor from the Phylip package with default parameters. Of these clades, only one member was chosen to represent the organism group in the final genome set. The 0.023 cutoff was chosen to ensure the exclusion of only highly related organisms because larger cutoffs resulted in sequence groups that were not monophyletic. A final set of 450 prokaryotic proteomes was used in this study (table S1).

Construction of a phylogenetic tree for prokaryotes

The following 12 ribosomal protein sequences were retrieved from the proteome set with hmmsearch of HMMER (*62*) and Pfam domain models with the Pfam gathering thresholds: L3 (Ribosomal_L3), L5 (Ribosomal_L5_C), L11 (Ribosomal_L11), L13 (Ribosomal_L13), L14 (Ribosomal_L14), S3 (Ribosomal_S3_C), S5 (Ribosomal_S5_C), S7 (Ribosomal_S7), S8 (Ribosomal_S8), S9 (Ribosomal_S9), S11 (Ribosomal_S11), and S17 (Ribosomal_S17). In rare instances where multiple copies of a protein are encoded in the same genome, the copies were found to be identical or virtually identical and only one was collected for further analysis. The ribosomal sequence sets were individually aligned in MAFFT. The alignments were concatenated, and poorly conserved positions in the alignment were eliminated by means of Gblocks with a maximum of eight contiguous nonconserved positions, a 10 length minimum for a block, allowed positions with a gap for <50% of the sequences, and a 50% + 1 minimum number of sequences for a conserved or flanking position. The resulting alignment was used to build a maximum likelihood tree in PhyML (Fig. 2).

Identification of chemotaxis proteins in genomic data sets

We identified CheA, CheB, CheD, CheR, CheV, CheW, and MCP sequences in genomes of interest using previously described domain queries (*10*) against the MiST database (*71*). CheC, CheX, and CheZ proteins were identified by PSI-BLAST queries using the experimentally defined sequences from *B. subtilis* (CheC) (*72*), *T. maritima* (CheX) (*21*), and *E. coli* (CheZ) (*73*), because current domain models of CheC, CheX, and CheZ do not capture the diversity of these proteins (*10*). PSI-BLAST queries for CheC and CheX components converged in 12 and 9 iterations, respectively, with default parameters (*E* value threshold of 0.001). The PSI-BLAST query for CheZ required a more stringent cutoff (*E* value threshold of 7×10^{-4}) to best distinguish them from other helical proteins identified in the searches, which lack the active site residues (convergence after six iterations).

CheC and CheX are similar to each other, as well as to FliY and FliM components of the bacterial flagellum (*72*), but these two chemotaxis proteins were distinguished from flagellar proteins by domain architecture differences (table S5). CheC and CheX proteins were aligned together in MAFFT, and the VISSA tool (*68*) was used to visualize their predicted secondary structures to distinguish between the two proteins, which have distinct structural differences despite their sequence similarity (*21*). These differences also correspond to distinct clades in a neighbor-joining tree built from the alignment with MEGA.

The high level of similarity between the CheY protein and the receiver domain of two-component system response regulators presents a challenge to CheY identification that was beyond the scope of this investigation. All sequences encoding receiver domains [HMMER (*62*) and Response_reg Pfam domain model with Pfam gathering thresholds] within three genes of identified chemotaxis proteins were collected, and domain architecture analysis was used to group them into HKI, CheY, or response regulator categories.

Proteins encoding HATPase_c domains that preceded receiver domains were defined as HKIs and removed from the classification analysis. The remaining sequences were aligned to the Response_reg domain model by hmmalign from HMMER. Sequences with <50 amino acids beyond the N- and C-terminal portions of the receiver domain model were classified as putative CheY sequences. The remaining sequences were classified as response regulator proteins, and their domain architectures were retrieved from the MiST database (*71*). Receiver domains fused to chemotaxis phosphatases (CheC, CheX, and CheZ) were considered putative CheY sequences because it is unclear whether such proteins are CheY proteins fused to their cognate phosphatases or whether these function as response regulator proteins with phosphatase activity toward other CheY proteins. CheY and other response regulators that are encoded immediately downstream from putative HKIs (defined here as a protein with a HATPase_c or HisKA family domain not classified as a CheA, MAC, or HKIII) were presumed to interact with the HKIs and were removed from the analysis. Although this method will miss putative CheY proteins that are not encoded near other chemotaxis components, most chemotaxis systems predicted to regulate motility had at least one CheY in their gene neighborhoods (table S1).

Downloaded from stke.sciencemag.org on July 6, 2010

Similar to CheY, collecting and assigning all MCP sequences to specific chemotaxis systems were beyond the scope of this investigation. We collected all sequences encoding an MCP signaling domain [HMMER (*62*) and MCPSignal Pfam domain model with Pfam gathering thresholds], but only those within three open reading frames of previously identified chemotaxis genes were assigned to length classes with the previously derived HMMs and cutoffs (*30*).

Gene neighborhoods were built from NCBI genome feature files (.gff) with custom Perl scripts. BLAST queries using undefined protein sequences encoded in chemotaxis gene neighborhoods occasionally identified divergent members of these protein families (see table S5 for details on domain architecture queries and their sensitivities).

Identification of motility components

The sequences of experimentally identified flagellar system components from *E. coli* (FlhA) and *Halobacterium* sp. NRC-1 (FlaI) and the PilU Tfp component from *P. aeruginosa* were used to identify homologs in PSI-BLAST (*67*) queries against our 450-proteome set.

There are no specific domain models for the FlhA, FlaI, and PilU protein sequences used in PSI-BLAST queries, and they all share similarity with components of systems that are not involved in motility. FlhA is a conserved member of the bacterial flagellum and has sequence similarity to a component of the type III secretion system (*74*). FlaI is a member of the archaeal flagellum (*75*) with similarity to Tfp and type II secretion components (*76*). PilU is involved in Tfp retraction (*77*) and is homologous to PilT (another Tfp retraction component), as well to other secretion nucleoside triphosphatases (NTPases) (*78*). A FlaI PSI-BLAST search converged after seven iterations (*E* value threshold of 1^{-50}). It identified >500 sequences, more than could be FlaI homologs given the 45 archaeal genomes in the data set. Proteins below the top 29 hits in the list were annotated as gspE (a type II secretion protein) and often found in genomes with another top hit that was presumed to be a true FlaI. The 29 top hits were thus classified as FlaI proteins predicted to be involved in flagellar motility (table S1).

FlhA PSI-BLAST searches converged after three iterations (*E* value threshold of 0.001). The sequences of the 273 significant hits were collected and aligned in MAFFT. A subsequent neighbor-joining tree built in MEGA revealed a clade of 49 sequences presumed to be parts of type III secretion systems (TTSSs). In-depth sequenced-based analysis of the bacterial flagella revealed that flagellar components are vertically inherited (*29*). In our FlhA tree, the clade we assigned to be true FlhA proteins showed distinct taxonomic-based clades consistent with the previous findings, whereas the clade of 49 sequences predicted to be TTSS components showed sporadic groupings characteristic of horizontal transfer, which is common in TTSSs (*79*). Furthermore, 15 of the TTSS clade members are from *Chlaymdia* and *Myxococcus* species that have never been shown to have flagella, and 32 of the remaining TTSS clade members have representatives in the FlhA clade. Thus, the 224 sequences in the "true" clade were classified as FlhA proteins predicted to be involved in flagellar motility (table S1).

PilU PSI-BLAST searches converged after nine iterations (*E* value threshold of 1^{-50}) and identified 886 putative homologs predicted to include many types of secretion NTPases. The sequences were collected and aligned in MAFFT. The alignment revealed that experimentally described PilT and PilU sequences from *P. aeruginosa* (*80*) and *Synechocystis* sp. PCC6803 (*51*) lack an insertion following the conserved NTPase modules (as did many other sequences annotated as PilT or PilU). The insertion contains a conserved four-cysteine motif. A neighbor-joining tree built in MEGA from the multiple sequence alignment showed that one clade contained 359 of the 396 sequences lacking the cysteine motif insertion. The 359 sequences were thus classified as PilT/PilU Tfp retraction enzymes potentially involved in motility (table S1).

Multiple sequence alignment and phylogenetic analyses

The primary chemotaxis sequence sets in this analysis (CheA, CheB, CheC, CheD, CheV, CheW, CheX, and CheZ) were individually aligned by MAFFT. Full-length CheA, CheB, and CheR sequences were first aligned by the e-ins-i algorithm of MAFFT given the diverse multidomain architecture of many of the members. CheC, CheD, CheV, CheW, CheX, and CheZ are typically single-domain proteins suited to the l-ins-i algorithm of MAFFT. However, members of the CheC, CheD, CheW, and CheX sets, which had duplications of their respective domains, were split into multiple sequences corresponding to each domain before alignment.

The core regions of the proteins were identified on the basis of sequence conservation in the alignment and structural data (*12*, *21*, *73*, *81*–*84*). For each sequence set, the core region was isolated and realigned by means of the l-ins-i algorithm of MAFFT. The core alignments were used to build maximum likelihood trees in PhyML (figs. S1, S4, and S7).

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/3/128/ra50/DC1

Table S1. Chemotaxis and motility components across 450 genomes.
Table S2. MAC systems.
Table S3. HKIII systems.
Table S4. Sensor domain diversity of two-component and chemotaxis systems.
Table S5. Domain architecture query sensitivities.
Fig. S1. Maximum likelihood trees built from CheA, CheB, and CheR multiple sequence alignments.
Fig. S2. Chemotaxis System Classification Step 1: Identifying coevolving CheA (A), CheB (B), and CheR (R) sequences.
Fig. S3. Chemotaxis System Classification Step 2: Defining classes using phylogenomic markers.
Fig. S4. Maximum likelihood tree built from a concatenated CheA-CheB-CheR multiple sequence alignment of CheA-CheB-CheR proteins.
Fig. S5. Chemotaxis System Classification Step 3: Assigning unclassified CheA (A), CheB (B), and CheR (R) sequences to chemotaxis system classes.
Fig. S6. Chemotaxis System Classification Step 4: Assigning CheC (C), CheD (D), CheV (V), CheW (W), CheX (X), and CheZ (Z) sequences to chemotaxis classes.
Fig. S7. Maximum likelihood trees built from individual CheC, CheD, CheV, CheW, and CheZ multiple sequence alignments.
Fig. S8. Multiple sequence alignment of the putative MAC1 methylation region.
Fig. S9. Multiple sequence alignment of the putative MAC2 methylation region.
Fig. S10. Multiple sequence alignment of the HKIII phosphotransfer and HATPase regions.
References

REFERENCES AND NOTES

1. L. E. Ulrich, E. V. Koonin, I. B. Zhulin, One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* **13**, 52–56 (2005).
2. A. M. Stock, V. L. Robinson, P. N. Goudreau, Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
3. I. B. Zhulin, A. N. Nikolskaya, M. Y. Galperin, Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. *J. Bacteriol.* **185**, 285–294 (2003).
4. M. N. Alekshun, S. B. Levy, T. R. Mealy, B. A. Seaton, J. F. Head, The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat. Struct. Biol.* **8**, 710–714 (2001).
5. J. R. Tuckerman, G. Gonzalez, E. H. Sousa, X. Wan, J. A. Saito, M. Alam, M. A. Gilles-Gonzalez, An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control. *Biochemistry* **48**, 9764–9774 (2009).
6. M. Y. Galperin, Structural classification of bacterial response regulators: Diversity of output domains and domain combinations. *J. Bacteriol.* **188**, 4169–4182 (2006).
7. J. L. Appleby, J. S. Parkinson, R. B. Bourret, Signal transduction via the multi-step phosphorelay: Not necessarily a road less traveled. *Cell* **86**, 845–848 (1996).
8. G. H. Wadhams, J. P. Armitage, Making sense of it all: Bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024–1037 (2004).
9. G. L. Hazelbauer, J. J. Falke, J. S. Parkinson, Bacterial chemoreceptors: High-performance signaling in networked arrays. *Trends Biochem. Sci.* **33**, 9–19 (2008).

Downloaded from stke.sciencemag.org on July 6, 2010

10. K. Wuichet, R. P. Alexander, I. B. Zhulin, Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. *Methods Enzymol.* **422**, 1–31 (2007).
11. E. C. Kofoid, J. S. Parkinson, Transmitter and receiver modules in bacterial signaling proteins. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4981–4985 (1988).
12. A. M. Bilwes, L. A. Alex, B. R. Crane, M. I. Simon, Structure of CheA, a signal-transducing histidine kinase. *Cell* **96**, 131–141 (1999).
13. H. Szurmant, G. W. Ordal, Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.* **68**, 301–319 (2004).
14. D. Bhaya, A. Takahashi, A. R. Grossman, Light regulation of type IV pilus-dependent motility by chemosensor-like elements in *Synechocystis* PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7540–7545 (2001).
15. D. R. Zusman, A. E. Scott, Z. Yang, J. R. Kirby, Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nat. Rev. Microbiol.* **5**, 862–872 (2007).
16. J. E. Berleman, C. E. Bauer, Involvement of a Che-like signal transduction cascade in regulating cyst cell development in *Rhodospirillum centenum*. *Mol. Microbiol.* **56**, 1457–1466 (2005).
17. J. R. Kirby, D. R. Zusman, Chemosensory regulation of developmental gene expression in *Myxococcus xanthus*. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2008–2013 (2003).
18. J. W. Hickman, D. F. Tifrea, C. S. Harwood, A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14422–14427 (2005).
19. A. N. Bible, B. B. Stephens, D. R. Ortega, Z. Xie, G. Alexandre, Function of a chemotaxis-like signal transduction pathway in modulating motility, cell clumping, and cell length in the alphaproteobacterium *Azospirillum brasilense*. *J. Bacteriol.* **190**, 6365–6375 (2008).
20. J. E. Berleman, C. E. Bauer, A *che*-like signal transduction cascade involved in controlling flagella biosynthesis in *Rhodospirillum centenum*. *Mol. Microbiol.* **55**, 1390–1402 (2005).
21. S. Y. Park, X. Chao, G. Gonzalez-Bonet, B. D. Beel, A. M. Bilwes, B. R. Crane, Structure and function of an unusual family of protein phosphatases: The bacterial chemotaxis proteins CheC and CheX. *Mol. Cell* **16**, 563–574 (2004).
22. C. V. Rao, G. D. Glekas, G. W. Ordal, The three adaptation systems of *Bacillus subtilis* chemotaxis. *Trends Microbiol.* **16**, 480–487 (2008).
23. Z. T. Güvener, D. F. Tifrea, C. S. Harwood, Two different *Pseudomonas aeruginosa* chemosensory signal transduction complexes localize to cell poles and form and remould in stationary phase. *Mol. Microbiol.* **61**, 106–118 (2006).
24. S. L. Porter, G. H. Wadhams, J. P. Armitage, *Rhodobacter sphaeroides*: Complexity in chemotactic signalling. *Trends Microbiol.* **16**, 251–260 (2008).
25. L. Dons, E. Eriksson, Y. Jin, M. E. Rottenberg, K. Kristensson, C. N. Larsen, J. Bresciani, J. E. Olsen, Role of flagellin and the two-component CheA/CheY system of *Listeria monocytogenes* in host cell invasion and virulence. *Infect. Immun.* **72**, 3237–3244 (2004).
26. K. Terry, S. M. Williams, L. Connolly, K. M. Ottemann, Chemotaxis plays multiple roles during *Helicobacter pylori* animal infection. *Infect. Immun.* **73**, 803–811 (2005).
27. S. I. Bibikov, A. C. Miller, K. K. Gosink, J. S. Parkinson, Methylation-independent aerotaxis mediated by the *Escherichia coli* Aer protein. *J. Bacteriol.* **186**, 3730–3737 (2004).
28. F. D. Ciccarelli, T. Doerks, C. von Mering, C. J. Creevey, B. Snel, P. Bork, Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287 (2006).
29. R. Liu, H. Ochman, Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7116–7121 (2007).
30. R. P. Alexander, I. B. Zhulin, Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2885–2890 (2007).
31. J. Tamames, G. Casari, C. Ouzounis, A. Valencia, Conserved clusters of functionally related genes in two bacterial genomes. *J. Mol. Evol.* **44**, 66–73 (1997).
32. C. S. Goh, A. A. Bogan, M. Joachimiak, D. Walther, F. E. Cohen, Co-evolution of proteins with their interaction partners. *J. Mol. Biol.* **299**, 283–293 (2000).
33. F. Pazos, D. Juan, J. M. Izarzugaza, E. Leon, A. Valencia, Prediction of protein interaction based on similarity of phylogenetic trees. *Methods Mol. Biol.* **484**, 523–535 (2008).
34. T. Dandekar, B. Snel, M. Huynen, P. Bork, Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).
35. J. G. Lawrence, Gene organization: Selection, selfishness, and serendipity. *Annu. Rev. Microbiol.* **57**, 419–440 (2003).
36. D. Bhaya, Light matters: Phototaxis and signal transduction in unicellular cyanobacteria. *Mol. Microbiol.* **53**, 745–754 (2004).
37. K. Terry, A. C. Go, K. M. Ottemann, Proteomic mapping of a suppressor of non-chemotactic cheW mutants reveals that *Helicobacter pylori* contains a new chemotaxis protein. *Mol. Microbiol.* **61**, 871–882 (2006).
38. A. Ferrández, A. C. Hawkins, D. T. Summerfield, C. S. Harwood, Cluster II che genes from *Pseudomonas aeruginosa* are required for an optimal chemotactic response. *J. Bacteriol.* **184**, 4374–4383 (2002).
39. K. K. Koretke, A. N. Lupas, P. V. Warren, M. Rosenberg, J. R. Brown, Evolution of two-component signal transduction. *Mol. Biol. Evol.* **17**, 1956–1970 (2000).
40. L. E. Ulrich, I. B. Zhulin, The MiST2 database: A comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res.* **38**, D401–D407 (2010).
41. D. Shiomi, I. B. Zhulin, M. Homma, I. Kawagishi, Dual recognition of the bacterial chemoreceptor by chemotaxis-specific domains of the CheR methyltransferase. *J. Biol. Chem.* **277**, 42325–42333 (2002).
42. A. Andreeva, D. Howorth, J. M. Chandonia, S. E. Brenner, T. J. Hubbard, C. Chothia, A. G. Murzin, Data growth and its impact on the SCOP database: New developments. *Nucleic Acids Res.* **36**, D419–D425 (2008).
43. P. Puigbò, Y. I. Wolf, E. V. Koonin, Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J. Biol.* **8**, 59 (2009).
44. W. F. Doolittle, O. Zhaxybayeva, On the origin of prokaryotic species. *Genome Res.* **19**, 744–756 (2009).
45. O. Zhaxybayeva, K. S. Swithers, P. Lapierre, G. P. Fournier, D. M. Bickhart, R. T. DeBoy, K. E. Nelson, C. L. Nesbø, W. F. Doolittle, J. P. Gogarten, K. M. Noll, On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5865–5870 (2009).
46. J. Adler, Chemotaxis in bacteria. *Science* **153**, 708–716 (1966).
47. M.-A. Jiménez-Pearson, I. Delany, V. Scarlato, D. Beier, Phosphate flow in the chemotactic response system of *Helicobacter pylori*. *Microbiology* **151**, 3299–3311 (2005).
48. Z. Yang, Y. Geng, D. Xu, H. B. Kaplan, W. Shi, A new set of chemotaxis homologues is essential for *Myxococcus xanthus* social motility. *Mol. Microbiol.* **30**, 1123–1130 (1998).
49. H. C. Berg, *Random Walks in Biology* (Princeton Univ. Press, Princeton, NJ, 1983).
50. D. B. Dusenbery, Spatial sensing of stimulus gradients can be superior to temporal sensing for free-swimming bacteria. *Biophys. J.* **74**, 2272–2277 (1998).
51. D. Bhaya, N. R. Bianco, D. Bryant, A. Grossman, Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC6803. *Mol. Microbiol.* **37**, 941–951 (2000).
52. K. Lipkow, Changing cellular location of CheZ predicted by molecular simulations. *PLoS Comput. Biol.* **2**, e39 (2006).
53. J. Stock, G. Kersulis, D. E. Koshland Jr., Neither methylating nor demethylating enzymes are required for bacterial chemotaxis. *Cell* **42**, 683–690 (1985).
54. R. Barak, M. Eisenbach, Chemotactic-like response of *Escherichia coli* cells lacking the known chemotaxis machinery but containing overexpressed CheY. *Mol. Microbiol.* **31**, 1125–1137 (1999).
55. P. S. Poole, J. P. Armitage, Motility response of *Rhodobacter sphaeroides* to chemotactic stimulation. *J. Bacteriol.* **170**, 5673–5679 (1988).
56. R. A. Goldstein, O. S. Soyer, Evolution of taxis responses in virtual bacteria: Non-adaptive dynamics. *PLoS Comput. Biol.* **4**, e1000084 (2008).
57. E. V. Koonin, Y. I. Wolf, Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
58. Z. Y. Jiang, C. E. Bauer, Analysis of a chemotaxis operon from *Rhodospirillum centenum*. *J. Bacteriol.* **179**, 5712–5719 (1997).
59. H. C. Vlamakis, J. R. Kirby, D. R. Zusman, The Che4 pathway of *Myxococcus xanthus* regulates type IV pilus-mediated motility. *Mol. Microbiol.* **52**, 1799–1811 (2004).
60. J. Adler, Bacterial chemotaxis and molecular neurobiology. *Cold Spring Harb. Symp. Quant. Biol.* **48**, 803–804 (1983).
61. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
62. S. R. Eddy, Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
63. K. Katoh, K. Kuma, H. Toh, T. Miyata, MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
64. K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
65. J. Felsenstein, PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166 (1989).
66. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
67. S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
68. L. E. Ulrich, I. B. Zhulin, Four-helix bundle: A ubiquitous sensory module in prokaryotic signal transduction. *Bioinformatics* **21** (Suppl. 3), iii45–iii48 (2005).
69. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).
70. M. Müller, A. Hildebrandt, Nucleotide sequences of the 23S rRNA genes from *Bordetella pertussis*, *B. parapertussis*, *B. bronchiseptica* and *B. avium*, and their implications for phylogenetic analysis. *Nucleic Acids Res.* **21**, 3320 (1993).
71. L. E. Ulrich, I. B. Zhulin, MiST: A microbial signal transduction database. *Nucleic Acids Res.* **35**, D386–D390 (2007).

Downloaded from stke.sciencemag.org on July 6, 2010

72. J. R. Kirby, C. J. Kristich, M. M. Saulmon, M. A. Zimmer, L. F. Garrity, I. B. Zhulin, G. W. Ordal, CheC is related to the family of flagellar switch proteins and acts independently from CheD to control chemotaxis in *Bacillus subtilis*. *Mol. Microbiol.* **42**, 573–585 (2001).
73. R. Zhao, E. J. Collins, R. B. Bourret, R. E. Silversmith, Structure and catalytic mechanism of the *E. coli* chemotaxis phosphatase CheZ. *Nat. Struct. Biol.* **9**, 570–575 (2002).
74. A. Blocker, K. Komoriya, S. Aizawa, Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3027–3030 (2003).
75. N. Patenge, A. Berendes, H. Engelhardt, S. C. Schuster, D. Oesterhelt, The *fla* gene cluster is involved in the biogenesis of flagella in *Halobacterium salinarum*. *Mol. Microbiol.* **41**, 653–663 (2001).
76. C. R. Peabody, Y. J. Chung, M. R. Yen, D. Vidal-Ingigliardi, A. P. Pugsley, M. H. Saier Jr., Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* **149**, 3051–3072 (2003).
77. P. Chiang, M. Habash, L. L. Burrows, Disparate subcellular localization patterns of *Pseudomonas aeruginosa* type IV pilus ATPases involved in twitching motility. *J. Bacteriol.* **187**, 829–839 (2005).
78. P. J. Planet, S. C. Kachlany, R. DeSalle, D. H. Figurski, Phylogeny of genes for secretion NTPases: Identification of the widespread *tadA* subfamily and development of a diagnostic key for gene classification. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2503–2508 (2001).
79. M. Naum, E. W. Brown, R. J. Mason-Gamer, Phylogenetic evidence for extensive horizontal gene transfer of type III secretion system genes among enterobacterial plant pathogens. *Microbiology* **155**, 3187–3199 (2009).
80. C. B. Whitchurch, J. S. Mattick, Characterization of a gene, *pilU*, required for twitching motility but not phage sensitivity in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **13**, 1079–1091 (1994).
81. X. Chao, T. J. Muff, S.-Y. Park, S. Zhang, A. M. Pollard, G. W. Ordal, A. M. Bilwes, B. R. Crane, A receptor-modifying deamidase in complex with a signaling phosphatase reveals reciprocal regulation. *Cell* **124**, 561–571 (2006).
82. G. S. Anand, P. N. Goudreau, A. M. Stock, Activation of methylesterase CheB: Evidence of a dual role for the regulatory domain. *Biochemistry* **37**, 14038–14047 (1998).
83. S. Djordjevic, P. N. Goudreau, Q. Xu, A. M. Stock, A. H. West, Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1381–1386 (1998).
84. I. J. Griswold, H. Zhou, M. Matison, R. V. Swanson, L. P. McIntosh, M. I. Simon, F. W. Dahlquist, The solution structure and interactions of CheW from *Thermotoga maritima*. *Nat. Struct. Biol.* **9**, 121–125 (2002).
85. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
86. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
87. **Acknowledgments:** We thank E. V. Koonin and J. S. Parkinson for comments on the manuscript and discussions and R. P. Alexander and L. E. Ulrich for technical assistance and discussions. **Funding:** This work was supported, in part, by NIH grant GM72285 (I.B.Z.) and by funds from the U.S. Department of Energy (DOE) BioEnergy Science Center, which is supported by the Office of Biological and Environmental Research in the DOE Office of Science. This research used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the DOE under contract no. DE-AC05-00OR22725. **Author contributions:** K.W. collected sequence data and performed computational work; K.W. and I.B.Z. designed experiments, analyzed results, and wrote the paper. **Competing interests:** The authors declare that they have no competing interests.

Submitted 4 November 2009
Accepted 3 June 2010
Final Publication 29 June 2010
10.1126/scisignal.2000724
Citation: K. Wuichet and I. B. Zhulin, Origins and diversification of a complex signal transduction system in prokaryotes. *Sci. Signal.* **3**, ra50 (2010).

Downloaded from stke.sciencemag.org on July 6, 2010