REVIEW

Two-Component Signal Transduction Systems of *Xanthomonas* spp.: A Lesson from Genomics

Wei Qian,^{1,2} Zhong-Ji Han,¹ and Chaozu He¹

¹State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, P. R. China; ²National Center for Plant Gene Research, Beijing 100101, P. R. China

Submitted 11 July 2007. Accepted 26 October 2007.

The two-component signal transduction systems (TCSTSs), consisting of a histidine kinase sensor (HK) and a response regulator (RR), are the dominant molecular mechanisms by which prokaryotes sense and respond to environmental stimuli. Genomes of *Xanthomonas* generally contain a large repertoire of TCSTS genes (approximately 92 to 121 for each genome), which encode diverse structural groups of HKs and RRs. Among them, although a core set of 70 TCSTS genes (about two-thirds in total) which accumulates point mutations with a slow rate are shared by these genomes, the other genes, especially hybrid HKs, experienced extensive genetic recombination, including genomic rearrangement, gene duplication, addition or deletion, and fusion or fission. The recombinations potentially promote the efficiency and complexity of TCSTSs in regulating gene expression. In addition, our analysis suggests that a co-evolutionary model, rather than a selfish operon model, is the major mechanism for the maintenance and microevolution of TCSTS genes in the genomes of *Xanthomonas*. Genomic annotation, secondary protein structure prediction, and comparative genomic analyses of TCSTS genes reviewed here provide insights into our understanding of signal networks in these important phytopathogenic bacteria.

Additional keywords: gene regulation.

The genus *Xanthomonas* consists of a group of rod-shaped, obligately aerobic, gram-negative bacteria (Goncalves and Rosato 2002). *Xanthomonas* spp. occupy diverse ecological niches, such as in soil, water, and human blood as well as plant tissues (Li et al. 1990; Swings and Civerolo 1993). Particularly, the majority of these bacteria are phytopathogens that usually are differentiated into pathogenic variants (pathovars) on the basis of host range (Alvarez 2000). For example, the species *Xanthomonas campestris* can be classified into over 100 pathovars; each has the ability to infect a specific plant taxon (Alvarez 2000). Interestingly, during the long-term interaction with their hosts, some bacteria of the genus *Xanthomonas*, especially *X. campestris* pv. *vesicatoria* the causal agent

of bacterial spot disease of solanaceous plants, and *X. oryzae* pv. *oryzae*, the causal agent of bacterial blight of rice, evolved a highly specific “gene-for-gene” relationship with cultivated plants (Flor 1971; Gurlebeck et al. 2006), because disease resistance will occur only if a plant cultivar encoding a specific resistance (*R*) gene interacts with a bacterial strain encoding a corresponding avirulence (*avr*) gene (Staskawicz et al. 2001). The fact that *Xanthomonas* spp. can adapt to so many heterogeneous environments suggests a long-standing notion that they may be equipped with elegant cellular signaling systems, which enable them to monitor environmental stimuli and then respond promptly by fine-tuning physiological pathways.

In prokaryotes, two-component signal transduction systems (TCSTSs) represent the dominant sense-response mechanisms to regulate a wide array of physiological pathways (Parkinson and Kofoed 1992; Stock et al. 2000). The total number of TCSTS proteins encoded by a bacterial genome, together with other signaling proteins, can be used as a measure of the adaptive potential of the organism (i.e., the bacterial intelligence quotient or “IQ”) (Galperin 2005). The prototype of TCSTS consists of a membrane-bound histidine kinase sensor (HK) and a cytoplasmic response regulator (RR). After detecting a specific stimulus, HK can be autophosphorylated on a conserved histidine residue within its transmitter domain, and then transfer the phosphoryl group onto a conserved aspartic acid residue within the N-terminal receiver domain of cognate RR. In due course, the activated RR causes adaptive changes by modulating expression of downstream genes or cellular machinery with its C-terminal output domain (Parkinson and Kofoed 1992; Stock et al. 2000). In addition, an atypical subfamily of HK which contains an additional phosphorylatable receiver domain has been identified and named “hybrid histidine kinase sensor” (HyHK). HyHK usually is involved in a four-step phosphorelay (His-Asp-His-Asp) via the histidine-containing phosphotransfer (Hpt) domain rather than the classic two-step phosphorelay (His-Asp) of TCSTS (Stock et al. 2000). In consequence, the amount, organization, protein secondary structure, and genomic dynamics of TCSTS largely define the ability of a bacterium to adapt to stochastic environmental changes.

Although TCSTSs play such an important role in regulating bacterial gene expression and cellular behavior, they have not been studied extensively in *Xanthomonas* spp. Only a few TCSTSs, including RpfC/RpfG (regulation of pathogenicity factors) (Tang et al. 1991), HrpG (hypersensitive response pathogenicity) (Wengelnik et al. 1996), and RaxH/RaxR (required for AvrXa21 activity) (Burdman et al. 2004), have been identified before. All the experimentally confirmed TCSTSs are associated with virulence: RpfC (a HyHK) and RpfG (a

Corresponding authors: Chaozu He; Telephone: 86-10-64807397; Fax: 86-10-64858245; E-mail: hecz@sun.im.ac.cn; and Wei Qian; Telephone: 86-10-64861838; Fax: 86-10-64858245; E-mail: qianw@sun.im.ac.cn

*The e-Xtra logo stands for “electronic extra” and indicates supplemental material is available online. Material published online includes four supplemental tables, one supplemental figure, and supplementary methods information.

RR) modulate the synthesis of extracellular enzymes, exopolysaccharides (EPS), and cell-to-cell signaling (Slater et al. 2000; Tang et al. 1991). Recently, the HD-GYP domain of RpfG was found to have the phosphodiesterase activity that degrades novel second messenger bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) (Ryan et al. 2006; Slater et al. 2000), providing a novel insight into the virulence regulation. Very recently, Dow and colleagues systematically characterized 37 genes encoding GGDEF, EAL, and HD-GYP domains involved in c-di-GMP signaling of *X. campestris* pv. *campestris*. Seven of these genes encode putative HKs or RRs (Ryan et al. 2007; this review). *hrpG* encodes a putative RR (Wengelnik et al. 1996). By interacting with downstream AraC-family transcription factor HrpX, HrpG controls the expression of the type III secretion system, the critical transportation machinery that delivers effectors into host cells (Noel et al. 2001). Although the cognate HK of HrpG remains unknown, a positive, GntR-family regulator Trh (transcriptional regulator for *hrp*) that is involved in expression of *hrpG* has been identified (Tsuge et al. 2006). In addition, evidence indicated that a MarR-family transcriptional factor HpaR (hypersensitive response and pathogenicity-associated regulator) might be a critical regulator of HrpG-dependent cascade (Noel et al. 2001; Qian et al. 2005; Wei et al. 2007). Identification of these proteins begins to uncover the signaling networks of HrpG regulon. In *X. oryzae* pv. *oryzae*, RaxH/RaxR affects the expression of a sulphation-related *raxSTAB* operon and have association with a race-specific *X. oryzae* pv. *oryzae*-rice recognition mediated by AvrXa21, a putative type I secretion system-dependent peptide (Burdman et al. 2004). Furthermore, RaxH/RaxR may be involved in regulation of the density-dependent expression of the rax genes (Lee et al. 2006). These findings suggested that RaxH/RaxR takes part in a novel regulatory mechanism of plant innate immunity

response. To date, however, a genome-scale study has not been carried out to explore the function of TCSTSs in *Xanthomonas* spp.

At present, six complete *Xanthomonas* genomes have been published, including two strains of *X. campestris* pv. *campestris* (da Silva et al. 2002; Qian et al. 2005), two strains of *X. oryzae* pv. *oryzae* (Lee et al. 2005; Ochiai 2005), a strain of *X. campestris* pv. *vesicatoria* (Thieme et al. 2005), and a strain of *X. axonopodis* pv. *citri* that causes citrus canker disease (da Silva et al. 2002). The sequencing projects have generated the largest dataset from a single genus of plant pathogens. Completion of so many closely related genomes provides researchers with a unique opportunity to not only survey TCSTSs on a genomic scale that will provide insight into their biological functions but also to analyze genetic factors directly contributing to microevolution of TCSTS genes. In this review, we re-annotated and analyzed all putative TCSTS genes in the genomes of *Xanthomonas* spp. by comparative genomic and evolutionary approaches (methods employed to analyze genomic data are provided as Supplementary Methods). The information may have immediate implication in future studies on mapping the cellular signaling networks in these important bacterial phytopathogens.

General features of TCSTS in genomes of *Xanthomonas* spp.

A large repertoire of TCSTSs was compiled from genomes of *Xanthomonas* spp. by a stringent method coupling similarity search and recognition of characteristic protein motifs. General features of the total of 632 predicted TCSTS genes (including orthodox HKs, HyHKs, and RRs) are listed in Table 1 (for details of these genes see Supplementary Table S1). Each *Xanthomonas* genome encodes numerous TCSTSs on its chromo-

Table 1. General features of two-component signal transduction systems (TCSTSs) in six genomes of *Xanthomonas*

Gene family ^b	General feature ^a					
	No. of genes	Average length (bp)	Genome (%)	GC content	GC3s	Nc
<i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC 33913						
Orthodox HK	32	1,495.2	0.94	0.657	0.832 ± 0.043	35.1
HyHK	20	2,846.7	1.12	0.656	0.835 ± 0.029	34.3
RR	54	787.9	0.82	0.647	0.830 ± 0.057	34.9
Total	106	...	2.88	0.662	0.831 ± 0.048	34.8
<i>X. campestris</i> pv. <i>campestris</i> 8004						
Orthodox HK	32	1,494.3	0.94	0.657	0.832 ± 0.043	35.2
HyHK	20	2,846.7	1.12	0.656	0.834 ± 0.028	34.3
RR	54	769.7	0.82	0.647	0.830 ± 0.057	34.8
Total	106	...	2.88	0.652	0.831 ± 0.049	34.8
<i>X. axonopodis</i> pv. <i>citri</i> 306						
Orthodox HK	35	1,490.4	1.01	0.653	0.808 ± 0.055	37.1
HyHK	21	2,755.3	1.12	0.644	0.808 ± 0.035	35.9
RR	58	769.3	0.86	0.640	0.807 ± 0.063	36.1
Total	114	...	2.99	0.644	0.807 ± 0.056	36.4
<i>X. campestris</i> pv. <i>vesicatoria</i> 85-10						
Orthodox HK	37	1,553.2	1.11	0.651	0.814 ± 0.049	36.4
HyHK	23	2,732.0	1.21	0.646	0.806 ± 0.028	35.9
RR	61	780.8	0.92	0.642	0.811 ± 0.048	35.7
Total	121	...	3.24	0.645	0.811 ± 0.045	35.9
<i>X. oryzae</i> pv. <i>oryzae</i> KACC10331						
Orthodox HK	28	1,451.9	0.79	0.642	0.799 ± 0.031	37.5
HyHK	12	2,709.0	0.66	0.636	0.776 ± 0.034	38.4
RR	52	901.1	0.93	0.632	0.783 ± 0.056	37.9
Total	92	...	2.38	0.636	0.787 ± 0.047	37.9
<i>X. oryzae</i> pv. <i>oryzae</i> MAFF 311018						
Orthodox HK	28	1,439.3	0.82	0.644	0.801 ± 0.033	37.2
HyHK	15	2,930.4	0.89	0.631	0.769 ± 0.048	38.7
RR	50	794.4	0.80	0.636	0.797 ± 0.045	36.6
Total	93	...	2.51	0.637	0.794 ± 0.043	37.1

^a GC3s = GC content in the third codon position, presented as average value. Standard deviation is calculated. Nc = effective number of codons.

^b HK = gene encoding histidine kinase sensor, HyHK = gene encoding hybrid histidine kinase sensor, and RR = gene encoding response regulator.

some (approximately 92 to 121 genes), but no TCSTS gene was found to be encoded by their plasmids. The total number of nucleotides from TCSTS genes takes up approximately 2.38 to 3.24% of the whole chromosome (Table 1). This result is according to the previous conclusion that bacteria distributed in diverse ecological niches tend to encode larger amount of TCSTSs than those living in limited or obligatory environments (Galperin 2005; Rodrigue et al. 2000). In addition, it reminds researchers that only a small portion of TCSTS genes of the genus *Xanthomonas* (<10%) has been experimentally documented; thus, our understanding of TCSTS regulation in these phytopathogenic bacteria needed to be enhanced by future research.

Two *X. campestris* pv. *campestris* genomes share a nearly identical set of TCSTSs in terms of both gene content and DNA sequences. For convenience, we only present the results from *X. campestris* pv. *campestris* ATCC 33913 in the following review, unless otherwise indicated. Unlike this, the two *X. oryzae* pv. *oryzae* genomes have variation in TCSTS gene contents. For example, XO_1378 of *X. oryzae* pv. *oryzae* MAFF 311018 was annotated as an HK. However, the protein product of its ortholog XO01477 from *X. oryzae* pv. *oryzae* KACC10331 does not encode an HK because it is truncated at the N-terminal region and lacks the most important phosphorylation sites (the H-box). A more complex example is found in the case of HyHK XO_1651. This gene has two homologs in *X. oryzae* pv. *oryzae* KACC10331 (i.e., XO01750 and XO01751). However, examination of the protein domain shows that XO01750 is not a TCSTS gene because its product contains only four Hpt domains and lacks any ATPase domain. Although XO01751 encodes receiver and ATPase domains, it

lacks recognizable phosphorylation sites (H-box) so that we annotated it as an RR rather than HyHK. Under the criteria employed in this review, 92 TCSTS genes in *X. oryzae* pv. *oryzae* KACC10331 and 93 in *X. oryzae* pv. *oryzae* MAFF 311018 (Table 1), in all, eventually were compiled, remarkably less than that of *X. campestris* pv. *campestris*, *X. campestris* pv. *vesicatoria*, and *X. axonopodis* pv. *citri*.

Furthermore, analyses of codon usage bias of TCSTSs showed that the protein coding strategy of *X. campestris* pv. *campestris*, *X. campestris* pv. *vesicatoria*, and *X. axonopodis* pv. *citri* are very similar (Table 1). However, two genomes of *X. oryzae* pv. *oryzae* have the lowest value of GC content in the third codon position (GC3s) and the highest value of effective number of codons (Nc, a parameter ranging from 20 to 61 to simply measure the extent of codon preference in a gene) (Wright 1990) than the other genomes (Table 1), suggesting that *X. oryzae* pv. *oryzae* could use more codons in translating TCSTS, and the pattern of codon usage of *X. oryzae* pv. *oryzae* is slightly biased from the other genomes.

Subcellular location and structural classification of TCSTS proteins.

Subcellular location. Subcellular location of all TCSTS proteins was predicted by PSORTb (Rey et al. 2005). In all, 64.7% of the annotated HKs (196 of 303 proteins, including HyHKs) are proteins having putative periplasmic sensor domains because they contain approximately 1 to 13 hydrophobic, transmembrane (TM) helices at the N-terminal sensor region. By contrast, 35.4% of HKs (107 proteins) lack a recognizable TM domain, which is similar to the chemotaxis kinase CheA and nitrogen regulator NtrB of *Escherichia coli* (Grebe and

Fig. 1. Structural classification of histidine kinases (HKs) based on phosphorylation sites (H-boxes). Each sequence logo is derived from multi-alignment of all proteins belonging to corresponding HK group and created by Weblogo. The height of each residue is according to the level of conservation at that site so that degree of conservation is illustrated. Numbers above each group indicate the numbers of HKs in the genomes of *Xanthomonas campestris* pv. *campestris* ATCC 33913, *X. campestris* pv. *campestris* 8004, *X. axonopodis* pv. *citri* 306, *X. oryzae* pv. *oryzae* MAFF 311018, *X. oryzae* pv. *oryzae* KACC10331, and *X. campestris* pv. *vesicatoria* 85-10, respectively.

Stock 1999). These HKs must be soluble, cytoplasmic proteins to sense intracellular stimuli. Meanwhile, all RRs were predicted to be located at cytosol.

Classification of HKs. Because biological functions of the majority of TCSTS in *Xanthomonas* spp. are unknown, structural classification of putative HKs and RRs will be a prerequisite for studying their regulatory mechanism. Among them, HKs are highly variable in length and in amino acid sequences, and their classification usually was based on alignment of H-boxes that contained phosphorylatable His residue, rather than by alignment of global protein sequences (Fabret et al. 1999; Grebe and Stock 1999). By applying the classification scheme of previous studies (Fabret et al. 1999; Kim and Forst 2001), multi-alignment using Clustal X classified all HKs into five distinct groups (Fig. 1). Group I HKs contain an invariant Pro residue at position five downstream from the invariant His. This group is the most common type of HK (66%) and can be further sorted into four subgroups according to alignment (group IA to ID). Group II HKs possess a conserved Asn at position four downstream from His. This group also contains most of the HyHKs. Group III HKs are characterized by an ERxxxR-E/D-L motif upstream from the His site of phosphorylation. Group IV HKs are distinguished from the others by a Pro preceding the conserved His, and a Phe instead of the usual acidic residue (Glu, Asp, or Asn) substituting at the position following the H-box histidine. As revealed by protein crystallization, group I to IV HKs usually have a similar structure to osmosensor EnvZ of *E. coli* (named as structural “class I HK”) (West and Stock 2001). Their H-boxes are immediately followed by C-terminal catalytic and ATP-binding domains, and serve as both a substrate for autophosphorylation and a dimerization region. Most of these HKs are multifunctional proteins that have kinase, phosphatase, and phosphotransferase functions. All these enzymatic activities require the phosphor-accepting His residue, and the enzymatic balance among them depends on signal input as well as collaboration between the H-box and other signaling domains (Dutta et al. 1999).

Distinct from other groups, the structure of group V HKs (approximately four to five proteins in each genome of the genus *Xanthomonas*) is similar to that of the chemotaxis sensor CheA (structural “class II HK”). The H-box His of group V is followed by a Ser or Thr rather than by the typical acidic residues, and the phosphorylation site is located at a domain similar to the Hpt domain of the complex phosphorelay systems (Fig. 1). As structural studies indicate (Bilwes et al. 1999), CheA lacks phosphatase activity, and its H-box is located in the so-called P1 domain (Hpt-like) at the N-terminus to act as a phosphorylation site. However, its dimerization function is carried out by a separate P3 domain rather than by the H-box (Bilwes et al. 1999; Mourey et al. 2001).

Classification of RRs. In contrast to HKs, RRs are highly conserved so that their classification usually is based on the structural similarity of C-terminal output domains (Stock et al. 2000). By referring to the criteria of Galperin (2006), all 329 annotated RRs were classified into diverse families by searching the pfam database (*e* value < 1.0; brief summary of the results are listed in Supplementary Table S2). This exercise showed, that except for the AraC and SpoIIE family RRs, *Xanthomonas* genomes encode almost all types of RRs commonly identified in other bacteria (Galperin 2006), indicating that their regulatory mechanisms are complicated and diverse.

Stand-alone RRs. Similar to the chemotaxis regulator, CheY, of *E. coli* (Baker et al. 2006) and sporulation regulator, Spo0F, of *Bacillus subtilis* (Feher et al. 1998), 93 RRs (28%) of *Xanthomonas* spp. lack any detectable output domain. These RRs may function by directly interacting with downstream proteins

rather than by functioning as transcription factors. Among them, each *Xanthomonas* genome has an RR that possesses two tandem receiver domains, in a pattern similar to the cell motility regulator, FrzZ, of the gram-positive bacterium *Myxococcus xanthus* (Inclan et al. 2007).

RRs with nucleic acid-binding domain. Approximately half of the RRs (172 proteins, 52%) contain a DNA-binding region as output domain. These RRs could be further classified into OmpR-, NarL-, NtrC-, PrrA-, and LytR-like proteins on the basis of the relatedness of their domain structure. Of these proteins, the OmpR (including the known RRs HrpG and RaxR) and NarL families are the most dominant types (127 proteins) in *Xanthomonas* spp. and both have a helix-turn-helix (HTH) motif at the C-terminal DNA-binding region. NtrC family RRs (27 proteins) often contain three domains: an N-terminal receiver domain, a C-terminal DNA-binding domain, and an unusual central output domain. Upon phosphorylation, the NtrC-family RRs may form oligomers and promote formation of open complexes via σ^{54} holoenzyme. Oligomerization, ATP hydrolysis, and interaction with σ^{54} factor all are functions of the central output domain (Lee et al. 2000). Each *Xanthomonas* genome encodes only one PrrA family RR. Examination of the crystal structure shows that the PrrA family RRs have three helix-bundle HTH DNA-binding domains (Nowak et al. 2006). In addition, we identified 12 LytR-family RRs from these genomes. The LytR family RR tends to be found in free-living and opportunistic bacteria that have a relatively large genome (>2 Mb). Although the three-dimensional structure of the LytR family protein is unclear, it was predicted to contain a novel DNA-binding domain different from any known related structures (Nikolskaya and Galperin 2002).

In contrast to the above DNA-binding RRs, the AmiR family RRs with ANTAR domains are unusual in that they are RNA-binding proteins (Shu and Zhulin 2002). Crystal structure of the representative protein, AmiR, shows that it contains a coiled-coil and three-helix bundle at the C-terminal region (O'Hara et al. 1999). Among the studied genomes, only *X. campestris* pvs. *campestris* and *vesicatoria* were found to encode ANTAR family RRs, and functions of these proteins remain to be investigated. Although there are corresponding orthologs in the *X. axonopodis* pv. *citri* and *X. oryzae* pv. *oryzae* genomes, these orthologs lack the characteristic receiver domains; therefore, they were not annotated as RRs.

RRs with protein binding domains. Each studied genome has one to two RRs with the protein-binding domain CheW. In *E. coli*, the CheW protein is one of the key components of the machinery of chemotaxis regulation. It may bind to the chemotaxis HK sensor, CheA, at a highly conserved chemoreceptor signaling region to form a soluble receptor-signaling complex that controls CheA kinase activity (Francis et al. 2004). Occasionally, bacteria contain RRs with other protein-binding domains, such as PAS, GAF, and TPR domains (Galperin 2006). However, we did not find any of these RR types in the genus *Xanthomonas* and these domains commonly are encoded by HKs (data not shown).

RRs with enzymatic output domains. This group of RRs, including those with CheB, GGDEF, EAL, and HD-GYP (including known RR RpfG) domains, are conserved among *Xanthomonas* genomes. Each *Xanthomonas* genome encodes two RRs, each with a CheB output domain that may have methylesterase activity (Stock et al. 2000). Furthermore, each of the *Xanthomonas* genomes has three RRs with HD-GYP domains, two RRs with EAL domains, an RR with a GGDEF domain, and an RR with both GGDEF and EAL domains. The GGDEF domain was shown to have c-di-GMP cyclase activity for biosynthesis of this cellular second messenger (Paul et al. 2004), whereas EAL and HD-GYP were proven

to have phosphodiesterase activity that catalyzes degradation of c-di-GMP (Ryan et al. 2006; Tamayo et al. 2005). These domains may be involved in turnover of c-di-GMP to coordinate its concentration in bacterial cells. As revealed by genomic analysis, *Xanthomonas* spp. do not encode any adenylate cyclase, so that whether cAMP is synthesized and plays a role in regulating gene expression is unclear (da Silva et al. 2002; Qian et al. 2005). However, as Ryan and associates (2007) have shown, c-di-GMP-related genes are involved in multiple physiological pathways, suggesting that c-di-GMP rather than cAMP plays a critical role in regulating gene expression in *Xanthomonas* spp.

Evolutionary conserved TCSTS genes in *Xanthomonas* spp.

Compiled TCSTS genes could be aggregated into clusters of orthologous groups (we designated them as “XanCOGs”) by similarity search (blastn, $e < 10^{-5}$) coupled with phylogenetic tree construction, which provides a basis for evolutionary analyses (Makarova and Koonin 2007). Among these XanCOGs, 70 (21, 8, and 41 HK, HyHK, and RR groups, respectively) are shared by all six analyzed genomes, including virulence-associated genes *hrpG*, *rpfC/rpfG*, and *raxH/raxR* (Supplementary Table S3). In order to assess the effect of point mutations upon the evolution of TCSTS genes, we calculated coefficients of DNA polymorphisms for the 70 XanCOGs, because they were not affected by extensive genetic recombination. TCSTS genes have a low value for nucleotide diversity ($\pi = 0.100$ on average) that was defined as the average number of nucleotide differences per site between two randomly chosen DNA sequences (Nei and Li 1979) (Fig. 2A). Out of the various TCSTS genes, the HyHKs maintain the highest level of genetic polymorphism, with an average π value for HyHKs (0.136) being higher than that for HKs (0.103) or RRs (0.091).

To assess the general evolutionary rates in these XanCOGs, we estimated the average rate of nonsynonymous substitution (K_a) versus synonymous substitution of nucleotides (K_s) by calculating pairwise K_a/K_s within each XanCOG. Frequencies of K_a/K_s distribution among HKs, HyHKs, and RRs are shown in Figure 2B. In general, the K_a/K_s ratio of TCSTS genes is much lower than 1.0 (0.406 on average), suggesting that the

fixation rate of amino acid change was reduced by purifying selection or functional constraint. However, the K_a/K_s ratios of five XanCOGs, including XanCOG3106 (HKs, $K_a/K_s = 3.01$), XanCOG0483 (HyHKs, 2.34), XanCOG2153 (HyHKs, 2.50), XanCOG0779 (RRs, 5.34), and XanCOG2703 (RRs, 3.88), are higher than 2.0, suggesting that these genes may have been subject to positive selection that favors the accumulation of adaptive genetic variations. Meanwhile, although HKs, HyHKs, and RRs have similar nonsynonymous substitution rates ($K_s \approx 0.35$), the K_a/K_s ratio of HyHK genes (0.848) is also clearly higher than that of HKs (0.365) and RRs (0.340), suggesting that HyHKs evolved more quickly. Interestingly, our results indicate that XanCOG1886 is absolutely conserved during evolution, because no nonsynonymous substitution was detected among its orthologs. It is worth noting here that XCC1886 is involved in multistress tolerance, as indicated by phenotypic characterization of its null mutant (W. Qian, Z.-J. Han, and C. He, unpublished data). Furthermore, the frequency distribution of polymorphism is measured by Tajima's D (Fig. 2C), a parameter that represents a comparison of nucleotide diversity estimated from the observed number of polymorphic sites against estimated nucleotide diversity from the allele frequency of the polymorphic sites (Tajima 1995). This coefficient could be used as a statistic for testing selective neutrality on coding sequences of TCSTS genes. All tests of Tajima's D are nonsignificant, and the values of the most shared XanCOGs are approximately 0.5 to 1.0 (Fig. 2C), suggesting that most of them are near selective neutral.

The above results indicated that, apart from several orthologous gene groups, the shared XanCOGs generally have low levels of DNA polymorphism and slow evolutionary rates, and most of them may not be relaxed from purifying selection. It suggests that accumulation of point mutations has not been the predominant force contributing to genetic variation of TCSTS genes in *Xanthomonas* spp. Accordingly, one might infer that the shared TCSTS genes conduct similar functions in these closely related *Xanthomonas* spp., as inferred by similarity search and operon organization. However, caution must be exercised because this expectation is based on predictions and does not represent experimentally verified biological fact. An

Fig. 2. Patterns of DNA polymorphism in two-component signal transduction system genes shared by six genomes of *Xanthomonas*. Frequency distribution among different protein families and average value of **A**, the estimations of nucleotide diversity (π , Phi); **B**, nonsynonymous substitution (K_a) versus synonymous substitution of nucleotides (K_s); and **C**, Tajima's D in each cluster of orthologous genes (XanCOGs) are depicted. HK = gene encoding orthodox histidine kinase, HyHK = gene encoding hybrid histidine kinase, and RR = gene encoding response regulator.

example of an exception (XanCOG1958 between two *X. campestris* pv. *campestris* strains in association of virulence regulation) has been observed in our study (W. Qian, Z.-J. Han, and C. He, *unpublished data*).

Genomic dynamics of TCSTS genes.

Interestingly, although there are 70 TCSTS XanCOGs conserved among genomes of *Xanthomonas* (approximately 60% in total), extensive genetic recombinations, including large-scale genome rearrangement, addition or deletion, gene fusion or fission, and duplication, have led to TCSTS variation among these genomes. These variations provide important clues to understanding the genetic differentiation of gene regulation among these bacteria.

Genomic organization and gene complement.

Genomic organization and orthologous relationship of TCSTS genes among the six *Xanthomonas* chromosomes are shown in Figure 3. These data indicate that TCSTS genes on the chromosomes of the two *X. oryzae* pv. *oryzae* strains are essentially colinear, and the same is true among TCSTSs of *X. campestris* pv. *campestris* ATCC 33913, *X. campestris* pv. *vesicatoria* 85-10, and *X. axonopodis* pv. *citri* 306, though genetic recombination occasionally has changed the relative position of some genes. The figure also shows that genomic organization of TCSTS genes in the two *X. oryzae* pv. *oryzae* strains is quite different from that of the others, suggesting that substantial lineage-specific rearrangements have occurred on the *X. oryzae* pv. *oryzae* genomes during their evolution. Interestingly, although *X. campestris* pv. *campestris* ATCC 33913 and *X. campestris* pv. *campestris* 8004 strains contain a nearly identical set of TCSTS genes, a genome-wide translocation across the replication axis was identified, and this rearrangement made the most TCSTS genes (86 pairs of genes) distribute symmetrically at mirror-image positions between two genomes. As our previous study has indicated, this dramatic genome-wide translocation may have been introduced by a replication-directed, homologous recombination between two IS1478-related genes (Qian et al. 2005).

Addition or deletion of TCSTS genes has occurred substantially in each of the genomes studied, resulting in gene loss

and the existence of specific genes (Supplementary Table S4). Among them, the two *X. oryzae* pv. *oryzae* genomes show more gene deletion for each of them, having lost 37 genes, particularly HyHKs. Consequently, the *X. oryzae* pv. *oryzae* genomes encode fewer HyHKs (13 and 15 proteins, respectively) than the other genomes (approximately 20 to 23 proteins). It is remarkable that the putative cognate HK and RR tend to be kept intact in genomes or lost together. Among 17 specific genes, there are four pairs of TCSTS genes (47.1%) organized in the same operon or clustered together whereas, among a total of 131 identified lost orthologs, 29 pairs (44.3%) of cognate HK and RR genes were found to be absent from different genomes. This tendency toward gene dynamics implies that cognate HK and RR genes have a very close co-evolutionary relationship.

Gene fusion or fission and duplication.

Close examination of gene sequences revealed that fusion or fission and duplication also contribute to genetic variation of TCSTS genes, especially HyHKs. For example, the N-terminal sequence of a HyHK, XAC3643, is very similar to that of XCC0562 or XCV3761 (HKs) whereas its C-terminal region is similar to XCC0563 or XCV3760 (RRs). Meanwhile, the ortholog of this gene is absent from the two *X. oryzae* pv. *oryzae* genomes. These results strongly suggest that the HyHK XAC3643 originated directly from a gene fusion event that combined cognate HK and RR, most probably by mutation of the stop codon and resulting in read-through of the open reading frame.

One of the significant characteristics of *Xanthomonas* genomes is that they all have a genomic segment (approximately 13 kb) containing approximately two to four tandem-repeated, closely related HyHKs (approximately 3,543 to 3,738 bp) (Fig. 4A). Domain organization of these HyHK proteins is very similar, because all of their N-terminal sensor regions contain a Y_Y_Y domain of unknown function and a number of repeated Reg_prop domains, whereas the C-terminal transmitter regions contain ATPase and receiver domains. The average value of nucleotide diversity (π) of these genes is 0.397 ± 0.013 , and the average value of pairwise Ka/Ks is 0.744 ($Ka = 0.534 \pm 0.205$, $Ks = 0.748 \pm 0.238$). The Ka/Ks value is less than 1.0

Fig. 3. Linear genomic comparison of two-component signal transduction system (TCSTS) genes in *Xanthomonas* spp. Red lines represent genes with orthologous relationship among genomes. Triangles with different colors depict different TCSTS genes and their direction of transcription.

but higher than average for the TCSTS genes (0.406). In addition, a slide windows analysis using both synonymous and nonsynonymous sites (Fig. 4B) indicated that DNA polymorphism of the 5' region (approximately 1 to 2,200 bp, encoding sensor region) of these genes is higher than that of the 3' region (approximately 2,400-bp to termination, encoding the transmitter and receiver region). The result suggests that the sensor region of these HyHKs evolved faster than the signal transmitter region after gene duplication, which is probably advantageous for the development of an ability to sense novel environmental stimuli on the basis of established signaling cascades.

To trace the evolutionary history of gene duplication, we retrieved all known homologs by searching GenBank, with the protein sequences of XCC2846, XCC2847, and XCC2848 as queries (blastp search, $e < 10^{-5}$, length of matched protein sequence $>80\%$ coverage). In all, 62 homologs from α - and γ -Proteobacteria, Acidobacteria, GNS bacteria, and Clostridiales were identified. The phylogenetic tree showed that HyHKs from Xanthomonadaceae were clustered into three phylogroups (PI to PIII). The most recent ancestor gene of these HyHKs was inferred to duplicate once (d-1) before species

differentiation of *Xanthomonas*. After that, a second duplication (d-2) appears to have taken place to yield the identified paralogs within genomes of *X. campestris* pv. *campestris*, *X. oryzae* pv. *oryzae*, and *X. campestris* pv. *vesicatoria*. However, because *X. axonopodis* pv. *citri* has paralogs only in phylogroups PI and PIII, and *X. campestris* pv. *vesicatoria* contains an additional gene copy in PIII, we speculate that an additional duplication event (d-3) may have produced these paralogs in PIII. Moreover, it is noticeable that the matched HyHKs of the closely-related species *Xylella fastidiosa* (belonging to Xanthomonadaceae) were present only in phylogroups PI. One possible explanation is that *X. fastidiosa* differentiated earlier before duplication (d-1) and the paralogs identified in *X. fastidiosa* Dixon strain might have arisen from a strain-specific duplication event (Supplementary Figure S1). However, considering that proteins of *X. fastidiosa* did not form a sibling group from those of *Xanthomonas*, and especially because *X. fastidiosa* is a fastidious, xylem-limited phytopathogen whose genome (2.68 Mb) has experienced large-scale genome reduction (Simpson et al. 2000), we tend to believe that the bacterial species *X. fastidiosa* also originated after duplication (d-1) and that its paralogs in clade PII

Fig. 4. Genomic organization and polymorphism of a cluster of similar hybrid histidine kinase (HyHK) genes in the genomes of *Xanthomonas*. **A**, Organization of hybrid histidine kinase sensors (HyHKs). Arrows represent HyHKs and their direction. Arrows with the same background of frames indicate they have orthologous or paralogous relationship. Lengths of HyHKs are shown above them. CDS numbers are according to those of the *Xanthomonas campestris* pv. *campestris* ATCC 33913, *X. campestris* pv. *campestris* 8004, *X. axonopodis* pv. *citri* 306, *X. campestris* pv. *vesicatoria* 85-10, *X. oryzae* pv. *oryzae* KACC10331, and *X. oryzae* pv. *oryzae* MAFF 311018, respectively. **B**, Slide window analysis of nucleotide diversity (π) using both synonymous and nonsynonymous sites of HyHKs. Window size was set as 50 with a step of 9 bp. Location of domain is according to the result of pfam database search.

and PIII may have existed previously but have been lost during genome reduction.

In summary, TCSTS genes in *Xanthomonas* spp. exhibit an evolutionary pattern in which a core set of genes is conserved, whereas accessory genes and genomic organization are highly flexible. Among them, it is interesting that HyHKs have special evolutionary characters compared with orthodox HKs and RRs. In general, HyHKs remain abundant in the genomes of *Xanthomonas* spp. (Table 1), have a higher level of DNA polymorphism (Fig. 2A) and faster evolutionary rate as revealed by *Ka/Ks* ratio (Fig. 2B), and experienced frequent gene fusion or fission and duplication. Taken together, it indicates that HyHKs are active during the microevolution of *Xanthomonas* spp. As previous studies have revealed, HyHKs usually take part in four-step phosphorelays (His-Asp-His-Asp) via an Hpt domain. This complex process potentially leads HyHKs to offer additional checkpoints for regulating the rate of protein phosphorylation, and also provides more junction sites for cross-talk with other cellular signaling cascades (Rodrigue et al. 2000; Stock et al. 2000). Although the biological functions of the most annotated HyHKs are unclear, the only experimentally verified HyHK in the genus *Xanthomonas*, RpfC, provides a good example of the fact that HyHKs are extensively versatile in function. By modulating the transcription of *rpfa-I* genes with cognate RpfG (an RR), RpfC regulates two different functions: it positively modulates the expression of virulence factors, such as extracellular enzymes and exopolysaccharides, and negatively regulates the biosynthesis of diffusible signal factor, a fatty acid (*cis*-11-methyl-2-dodecenoic acid) that functions as a signaling molecule during cell-to-cell signaling of *Xanthomonas* spp., by its receiver domain (He et al. 2006). Furthermore, Ryan and associates (2006) confirmed the role of RpfC/RpfG in the degradation of cellular second messenger c-di-GMP directly, and functional genomic analysis has shown that RpfC/RpfG positively influences the transcription of three other c-di-GMP associated genes involved in *X. campestris* pv. *campestris* virulence (Ryan et al. 2007). Thus, the versatility of RpfC provides good evidence that other HyHKs of *Xanthomonas* spp. also may be involved in complex signaling networks that would be worth investigating.

Mechanism of maintenance of TCSTS in *Xanthomonas* spp.

It is worth noting that co-evolutionary relationships between cognate HK and RR have been detected previously by phylogenetic and genomic analyses (Fabret et al. 1999; Koretke et al. 2000). The present study provides additional evidence for the co-evolution by showing a remarkable feature: HK and RR that organize into an operon or in close proximity usually are deleted or acquired together. Co-evolution of TCSTSs by means of addition or deletion as a unit is likely to affect the entire regulons that they controlled, which may result immediately in alteration of bacterial fitness during environmental change. In turn, this would determine the direction of bacterial evolution. In general, co-evolution of genes organized in operons is explained by two major evolutionary models (i.e., the selfish operon model [SOM] and the co-regulation model [CRM]). The SOM proposes that operons can be formed by the gradual, stepwise formation of gene clusters via horizontal gene transfer (Lawrence and Roth 1996). During this process, intervening genes with unrelated functions are deleted gradually by selection or genetic drift without reducing the fitness of recipient bacterial cells, leading to genes coding for a single metabolic function being situated in close proximity. Once co-transcription under a single promoter has been established by progressively reinforced selection, the novel operon eventually evolves or is deleted efficiently and “selfishly” as an inte-

grated unit (Lawrence 1997). The SOM provides a selection mechanism and an intermediate step to the origin and maintenance of operons. It has been applied successfully to explain operon evolution, such as in the cases of the restriction modification system and the toxin-antitoxin system (Lawrence 1999; Naito et al. 1998). However, when considering our case, in which a short time scale of microevolution has occurred in closely related genomes of *Xanthomonas*, there are two opposing reasons that suggest that the SOM is not entirely appropriate in our case. i) The SOM predicts that essential genes should not cluster because they cannot undergo multiple rounds of gene gain or loss. However, although more direct genetic evidence is needed, *XCC2361*, *XCC2695*, and *XCC3643* may be essential genes because they could not be mutated by insertion inactivation (W. Qian, Z.-J. Han, and C. He, *unpublished data*). All of these genes are located in operons or in the vicinity of putative cognate HK genes (*XCC2360*, *XCC2694*, and *XCC3642*, respectively). Similarly, investigation of the genome of *E. coli* also has revealed that essential genes have an especially strong tendency to cluster compared with nonessential genes (Price et al. 2005). ii) Approximately half of the “orphan” TCSTS genes are absent in genomes, whereas their putative cognations remain present, suggesting that dynamic modes other than the SOM have occurred during TCSTS evolution.

We suggest the classical CRM (Jacob and Monod 1961) as the major mechanism to explain the maintenance of cognate TCSTS genes during the microevolution of *Xanthomonas* spp. The CRM proposes that genes are clustered into operons because co-regulation at a single promoter is selectively beneficial, and operons are under strong purifying selection because their disruption would destroy the co-regulatory relationship (Ermolaeva et al. 2001). In the case of TCSTS, a high level of substrate specificity generally is maintained between cognate HK and RR to ensure the precision of signal transduction. As studies on phosphotransfer kinetics of KinA/Spo0F and VanS/VanR have shown (Fisher et al. 1996; Grimshaw et al. 1998), HKs have a significant kinetic preference for their cognate RRs, with a magnitude approximately 10^3 to 10^4 times higher than other unrelated RRs. Based on this kinetic preference, a robust in vitro biochemical technique (phosphotransfer profiling) recently has been developed to screen unidentified HK/RR couples (Biondi et al. 2006; Skerker et al. 2005). Because the equilibrium of phosphorylated and dephosphorylated forms of RR is supposed to be central to TCSTS regulation (Stock et al. 2000), co-regulation with its cognate HK under a single promoter is selectively beneficial. As a consequence, unless the intact TCSTS has been recovered during evolution, the absence of one of its component genes eventually will result in deletion of another component by genetic drift or purifying selection, because the connection of the entire regulon has been disrupted and this will decrease bacterial fitness. Thus, although we could not exclude the possibility that some TCSTS genes can be gained or lost by a batch mode as described in SOM, the presence of orphan genes in the genomes of *Xanthomonas* spp. strongly supports a two-step addition or deletion mode for maintenance of TCSTS genes.

Conclusions and perspective.

The present comparative genomic survey showed that *Xanthomonas* spp. encode a large number of TCSTSs that are diverse in protein structure and subjected to extensive genetic recombination, suggesting that these proteins may be critical for survival in heterogeneous niches, and that the bacterial IQ scores (Galperin 2005) of *Xanthomonas* spp. are remarkably high. Recent progress has provided evidence that TCSTS involved in subtle regulation of virulence factors (Dow et al.

2003; He et al. 2006; Lee et al. 2006; Ryan et al. 2006). However, the physiological functions of most of the TCSTSs remain to be determined, although their phosphotransfer mechanisms can be predicted by secondary structure prediction. Therefore, reverse-genetic analysis is needed to identify phenotypic changes of null mutants, which will set up a solid basis for further research on gene function. In addition to identifying novel TCSTS genes directly regulating the expression of virulence factors, attention should be paid to TCSTSs that are important for nutrients uptake, energy metabolism, and degradation of harmful chemicals when bacteria grow in host tissues, because these genes may play important roles for survival and proliferation in planta.

TCSTS genes with unusual evolutionary history, including *HyHKs*, strain-specific genes, and genes that experienced addition or deletion or positive selection, are worthy to be specially noticed in a future study because rapid evolutionary changes in these genes may reflect reprogramming or reorganization of regulatory cascades during bacterial speciation. For example, *HKs* belonging to XanCOG3106 might be subject to positive selection because nonsynonymous substitutions are remarkably higher than synonymous substitutions ($Ka/Ks > 1.0$). The N-terminal sensor regions of these HKs evolved very quickly, whereas the C-terminal transmitter regions as well as their cognate RRs (XanCOG3107) remain relatively conserved (data not shown). This suggests that the phosphotransfer pathway and downstream regulated genes of XanCOG3106/XanCOG3107 regulon maintained stable during evolution, but the property of environmental stimuli sensing by XanCOG3106 may be different among *X. campestris* pv. *campestris*, *X. campestris* pv. *vesicatoria*, *X. oryzae* pv. *oryzae*, and *X. axonopodis* pv. *citri*. What are these stimuli? How do HKs detect them? Do these stimuli vary with different host plants? Answering these questions will not only help us to understand the regulatory mechanism but also provide a molecular genetic explanation for natural selection during adaptive evolution of pathogen.

X. oryzae pv. *oryzae* genomes, which have smaller genome sizes among *Xanthomonas* spp. (Lee et al. 2005; Ochiai 2005), experienced extensive gene loss (discarded approximately 15% TCSTS genes) and contain a smaller set of TCSTS genes (especially *HyHKs*) (Table 1). In addition, the genomic organization and codon usage bias of TCSTS genes of *X. oryzae* pv. *oryzae* are different from the other genomes. For example, deletion of a 25-kb genomic segment in genomes of *X. oryzae* pv. *oryzae* resulted in the absence of approximately 20 genes, including six XanCOGs (XanCOG1175, XanCOG1176, XanCOG1182, XanCOG1185, XanCOG1186, and XanCOG1187). Among them, XanCOG1175 and XanCOG1176 may regulate the expression of regulators of a σ^B factor, and XanCOG1182 and XanCOG1185 may regulate several chemotaxis-related proteins. Loss of σ factors and their regulatory elements (including TCSTS) also is observed in other pathogenic bacteria, such as *Mycobacterium leprae*, the causative agent of leprosy (Madan Babu 2003; Moran 2002; Tyagi and Saini 2004). It has been revealed that loss of these important signaling genes triggers the accumulation of mutations in downstream-regulated genes and creates a large set of pseudogenes (Cole et al. 2001; Tyagi and Saini 2004). Therefore, the lineage-specific elimination of TCSTS genes in *X. oryzae* pv. *oryzae* indicates that the corresponding regulons are important to other *Xanthomonas* spp. but unneeded to the lifestyle of *X. oryzae* pv. *oryzae*. Was this recently accelerated evolution of TCSTS of *X. oryzae* pv. *oryzae* driven by domestication of cultivated rice? How many regulatory differences are there between *X. oryzae* pv. *oryzae* and other *Xanthomonas* spp.? What roles does TCSTS regulation play in the specific gene-for-gene relationship between *X.*

oryzae pv. *oryzae* and rice cultivars? The availability of genomes of *Xanthomonas* spp. will facilitate future research efforts to try to answer these questions.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grants 30270027 and 30470036 to W. Qian) and the Ministry of Science and Technology of China (grant 2006AA02Z180 to C. He).

LITERATURE CITED

- Alvarez, A. M. 2000. Black rot of crucifers. Pages 21-52 in: Mechanisms of Resistance to Plant Diseases. A. J. Slusarenko, R. S. S. Fraser, and L. C. van Loon, eds. Kluwer Academic Publications, Dordrecht, The Netherlands.
- Baker, M. D., Wolanin, P. M., and Stock, J. B. 2006. Signal transduction in bacterial chemotaxis. *Bioessays* 28:9-22.
- Bilwes, A. M., Alex, L. A., Crane, B. R., and Simon, M. I. 1999. Structure of CheA, a signal-transducing histidine kinase. *Cell* 96:131-141.
- Biondi, E. G., Reisinger, S. J., Skerker, J. M., Arif, M., Perchuk, B. S., Ryan, K. R., and Laub, M. T. 2006. Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* 444:899-904.
- Burdman, S., Shen, Y., Lee, S. W., Xue, Q., and Ronald, P. 2004. RaxH/RaxR: a two-component regulatory system in *Xanthomonas oryzae* pv. *oryzae* required for AvrXa21 activity. *Mol. Plant-Microbe Interact.* 17:602-612.
- Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., MacLean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R., and Barrell, B. G. 2001. Massive gene decay in the leprosy bacillus. *Nature* 409:1007-1011.
- da Silva, A. C., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., Monteiro-Vitorello, C. B., Van Sluys, M. A., Almeida, N. F., Alves, L. M., do Amaral, A. M., Bertolini, M. C., Camargo, L. E., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L. P., Cicarelli, R. M., Coutinho, L. L., Cursino-Santos, J. R., El-Dorry, H., Faria, J. B., Ferreira, A. J., Ferreira, R. C., Ferro, M. I., Formighieri, E. F., Franco, M. C., Greggio, C. C., Gruber, A., Katsuyama, A. M., Kishi, L. T., Leite, R. P., Lemos, E. G., Lemos, M. V., Locali, E. C., Machado, M. A., Madeira, A. M., Martinez-Rossi, N. M., Martins, E. C., Meidanis, J., Menck, C. F., Miyaki, C. Y., Moon, D. H., Moreira, L. M., Novo, M. T., Okura, V. K., Oliveira, M. C., Oliveira, V. R., Pereira, H. A., Rossi, A., Sena, J. A., Silva, C., de Souza, R. F., Spinola, L. A., Takita, M. A., Tamura, R. E., Teixeira, E. C., Tezza, R. I., Trindade dos Santos, M., Truffi, D., Tsai, S. M., White, F. F., Setubal, J. C., and Kitajima, J. P. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417:459-463.
- Dow, J. M., Crossman, L., Findlay, K., He, Y. Q., Feng, J. X., and Tang, J. L. 2003. Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc. Natl. Acad. Sci. U.S.A.* 100:10995-11000.
- Dutta, R., Qin, L., and Inouye, M. 1999. Histidine kinases: diversity of domain organization. *Mol. Microbiol.* 34:633-640.
- Ermolaeva, M. D., White, O., and Salzberg, S. L. 2001. Prediction of operons in microbial genomes. *Nucleic Acids Res.* 29:1216-1221.
- Fabret, C., Feher, V. A., and Hoch, J. A. 1999. Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. *J. Bacteriol.* 181:1975-1983.
- Feher, V. A., Tzeng, Y. L., Hoch, J. A., and Cavanagh, J. 1998. Identification of communication networks in Spo0F: a model for phosphorylation-induced conformational change and implications for activation of multiple domain bacterial response regulators. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 425:1-6.
- Fisher, S. L., Kim, S. K., Wanner, B. L., and Walsh, C. T. 1996. Kinetic comparison of the specificity of the vancomycin resistance VanS for two response regulators, VanR and PhoB. *Biochemistry* 35:4732-4740.
- Flor, H. H. 1971. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9:275-296.
- Francis, N. R., Wolanin, P. M., Stock, J. B., Derosier, D. J., and Thomas, D. R. 2004. Three-dimensional structure and organization of a receptor/signaling complex. *Proc. Natl. Acad. Sci. U.S.A.* 101:17480-17485.
- Galperin, M. Y. 2005. A census of membrane-bound and intracellular sig-

- nal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol.* 5:35.
- Galperin, M. Y. 2006. Structural classification of bacterial response regulators: Diversity of output domains and domain combinations. *J. Bacteriol.* 188:4169-4182.
- Goncalves, E. R., and Rosato, Y. B. 2002. Phylogenetic analysis of *Xanthomonas* species based upon 16S-23S rDNA intergenic spacer sequences. *Int. J. Syst. Evol. Microbiol.* 52:355-361.
- Grebe, T. W., and Stock, J. B. 1999. The histidine protein kinase superfamily. *Adv. Microb. Physiol.* 41:139-227.
- Grimshaw, C. E., Huang, S., Hanstein, C. G., Strauch, M. A., Burbulys, D., Wang, L., Hoch, J. A., and Whiteley, J. M. 1998. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in *Bacillus subtilis*. *Biochemistry* 37:1365-1375.
- Gurlebeck, D., Thieme, F., and Bonas, U. 2006. Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. *J. Plant Physiol.* 163:233-255.
- He, Y. W., Wang, C., Zhou, L., Song, H., Dow, J. M., and Zhang, L. H. 2006. Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *J. Biol. Chem.* 281:33414-33421.
- Inclan, Y. F., Vlamakis, H. C., and Zusman, D. R. 2007. FrzZ, a dual CheY-like response regulator, functions as an output for the Frz chemosensory pathway of *Myxococcus xanthus*. *Mol. Microbiol.* 65:90-102.
- Jacob, F., and Monod, J. 1961. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* 26:193-211.
- Kim, D., and Forst, S. 2001. Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* 147:1197-1212.
- Koretke, K. K., Lupas, A. N., Warren, P. V., Rosenberg, M., and Brown, J. R. 2000. Evolution of two-component signal transduction. *Mol. Biol. Evol.* 17:1956-1970.
- Lawrence, J. 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr. Opin. Genet. Dev.* 9:642-648.
- Lawrence, J. G. 1997. Selfish operons and speciation by gene transfer. *Trends Microbiol.* 5:355-359.
- Lawrence, J. G., and Roth, J. R. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843-1860.
- Lee, B. M., Park, Y. J., Park, D. S., Kang, H. W., Kim, J. G., Song, E. S., Park, I. C., Yoon, U. H., Hahn, J. H., Koo, B. S., Lee, G. B., Kim, H., Park, H. S., Yoon, K. O., Kim, J. H., Jung, C. H., Koh, N. H., Seo, J. S., and Go, S. J. 2005. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Res.* 33:577-586.
- Lee, J., Owens, J. T., Hwang, I., Meares, C., and Kustu, S. 2000. Phosphorylation-induced signal propagation in the response regulator *ntnC*. *J. Bacteriol.* 182:5188-5195.
- Lee, S. W., Han, S. W., Bartley, L. E., and Ronald, P. C. 2006. From the Academy: Colloquium review. Unique characteristics of *Xanthomonas oryzae* pv. *oryzae* AvrXa21 and implications for plant innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* 103:18395-18400.
- Li, Z. X., Bian, Z. S., Zheng, H. P., Yue, Y. S., Yao, J. Y., Gong, Y. P., Cai, M. Y., and Dong, X. Z. 1990. First isolation of *Xanthomonas campestris* from the blood of a Chinese woman. *Chin. Med. J. (Taipei)* 103:435-439. (In English)
- Madan Babu, M. 2003. Did the loss of sigma factors initiate pseudogene accumulation in *M. leprae*? *Trends Microbiol.* 11:59-61.
- Makarova, K. S., and Koonin, E. V. 2007. Evolutionary genomics of lactic acid bacteria. *J. Bacteriol.* 189:1199-1208.
- Moran, N. A. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583-586.
- Mourey, L., Da Re, S., Pedelacq, J. D., Tolstykh, T., Faurie, C., Guillet, V., Stock, J. B., and Samama, J. P. 2001. Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. *J. Biol. Chem.* 276:31074-31082.
- Naito, Y., Naito, T., and Kobayashi, I. 1998. Selfish restriction modification genes: resistance of a resident R/M plasmid to displacement by an incompatible plasmid mediated by host killing. *Biol. Chem.* 379:429-436.
- Nei, M., and Li, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 76:5269-5273.
- Nikolskaya, A. N., and Galperin, M. Y. 2002. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* 30:2453-2459.
- Noel, L., Thieme, F., Nennstiel, D., and Bonas, U. 2001. cDNA-AFLP analysis unravels a genome-wide *hrpG*-regulon in the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Microbiol.* 41:1271-1281.
- Nowak, E., Panjikar, S., Konarev, P., Svergun, D. I., and Tucker, P. A. 2006. The structural basis of signal transduction for the response regulator PrrA from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 281:9659-9666.
- Ochiai, H. I., Takeya, Y., Sasaki, M., and Kaku, A. H. 2005. Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large number of effector genes and insertion sequences to its race diversity. *Jpn. Agric. Rep. Q.* 39:275-285.
- O'Hara, B. P., Norman, R. A., Wan, P. T., Roe, S. M., Barrett, T. E., Drew, R. E., and Pearl, L. H. 1999. Crystal structure and induction mechanism of AmiC-AmiR: a ligand-regulated transcription antitermination complex. *EMBO (Eur. Mol. Biol. Organ.) J.* 18:5175-5186.
- Parkinson, J. S., and Kofoid, E. C. 1992. Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.* 26:71-112.
- Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B., and Jenal, U. 2004. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* 18:715-727.
- Price, M. N., Huang, K. H., Arkin, A. P., and Alm, E. J. 2005. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res.* 15:809-819.
- Qian, W., Jia, Y., Ren, S. X., He, Y. Q., Feng, J. X., Lu, L. F., Sun, Q., Ying, G., Tang, D. J., Tang, H., Wu, W., Hao, P., Wang, L., Jiang, B. L., Zeng, S., Gu, W. Y., Lu, G., Rong, L., Tian, Y., Yao, Z., Fu, G., Chen, B., Fang, R., Qiang, B., Chen, Z., Zhao, G. P., Tang, J. L., and He, C. 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Res.* 15:757-767.
- Rey, S., Acab, M., Gardy, J. L., Laird, M. R., deFays, K., Lambert, C., and Brinkman, F. S. 2005. PSORTdb: a protein subcellular localization database for bacteria. *Nucleic Acids Res.* 33:D164-168.
- Rodrigue, A., Quentin, Y., Lazdunski, A., Mejean, V., and Foglino, M. 2000. Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol.* 8:498-504.
- Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y. W., Zhang, L. H., Heeb, S., Camara, M., Williams, P., and Dow, J. M. 2006. Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci. U.S.A.* 103:6712-6717.
- Ryan, R. P., Fouhy, Y., Lucey, J. F., Jiang, B. L., He, Y. Q., Feng, J. X., Tang, J. L., and Dow, J. M. 2007. Cyclic di-GMP signaling in the virulence and environmental adaptation of *Xanthomonas campestris*. *Mol. Microbiol.* 63:429-442.
- Shu, C. J., and Zhulin, I. B. 2002. ANTAR: an RNA-binding domain in transcription antitermination regulatory proteins. *Trends Biochem. Sci.* 27:3-5.
- Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bove, J. M., Briones, M. R., Bueno, M. R., Camargo, A. A., Camargo, L. E., Carraro, D. M., Carrer, H., Colauto, N. B., Colombo, C., Costa, F. F., Costa, M. C., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A. P., Ferreira, A. J., Ferreira, V. C., Ferro, J. A., Fraga, J. S., Franca, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuramae, E. E., Laigret, F., Lambais, M. R., Leite, L. C., Lemos, E. G., Lemos, M. V., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M., Madeira, H. M., Marino, C. L., Marques, M. V., Martins, E. A., Martins, E. M., Matsukuma, A. Y., Menck, C. F., Miracca, E. C., Miyaki, C. Y., Montero-Vitello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L., Netto, L. E., Nhani, A., Jr., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A., Pereira, H. A., Jr., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M. Rosa, A. J., de Rosa, V. E., Jr., de Sa, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C., da Silva, A. M., da Silva, F. R., da Silva, W. A., Jr., da Silveira, J. F., Silvestri, M. L., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tshako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J., and Setubal, J. C. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406:151-159.
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., and Laub, M. T. 2005. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol.* 3:e334.
- Slater, H., Alvarez-Morales, A., Barber, C. E., Daniels, M. J., and Dow, J. M. 2000. A two-component system involving an HD-GYP domain protein links cell-cell signaling to pathogenicity gene expression in *Xanthomonas campestris*. *Mol. Microbiol.* 38:986-1003.
- Staskawicz, B. J., Mudgett, M. B., Dangl, J. L., and Galan, J. E. 2001. Common and contrasting themes of plant and animal diseases. *Science* 292:2285-2289.

- Stock, A. M., Robinson, V. L., and Goudreau, P. N. 2000. Two-component signal transduction. *Annu. Rev. Biochem.* 69:183-215.
- Swings, J. G., and Civerolo, E. L. 1993. *Xanthomonas*. Chapman & Hall, London.
- Tajima, F. 1995. Effect of non-random sampling on the estimation of parameters in population genetics. *Genet. Res.* 66:267-276.
- Tamayo, R., Tischler, A. D., and Camilli, A. 2005. The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *J. Biol. Chem.* 280:33324-33330.
- Tang, J. L., Liu, Y. N., Barber, C. E., Dow, J. M., Wootton, J. C., and Daniels, M. J. 1991. Genetic and molecular analysis of a cluster of *rpf* genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *campestris*. *Mol. Gen. Genet.* 226:409-417.
- Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Buttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., Kirchner, O., Lanz, C., Linke, B., McHardy, A. C., Meyer, F., Mittenhuber, G., Nies, D. H., Niesbach-Klosgen, U., Patschkowski, T., Ruckert, C., Rupp, O., Schneiker, S., Schuster, S. C., Vorholter, F. J., Weber, E., Puhler, A., Bonas, U., Bartels, D., and Kaiser, O. 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* 187:7254-7266.
- Tsuge, S., Nakayama, T., Terashima, S., Ochiai, H., Furutani, A., Oku, T., Tsuno, K., Kubo, Y., and Kaku, H. 2006. Gene involved in transcriptional activation of the *hrp* regulatory gene *hrpG* in *Xanthomonas oryzae* pv. *oryzae*. *J. Bacteriol.* 188:4158-4162.
- Tyagi, J. S., and Saini, D. K. 2004. Did the loss of two-component systems initiate pseudogene accumulation in *Mycobacterium leprae*? *Microbiology* 150:4-7.
- Wei, K., Tang, D. J., He, Y. Q., Feng, J. X., Jiang, B. L., Lu, G. T., Chen, B., and Tang, J. L. 2007. *hpaR*, a putative *marR* family transcriptional regulator, is positively controlled by HrpG and HrpX and involved in the pathogenesis, hypersensitive response, and extracellular protease production of *Xanthomonas campestris* pathovar *campestris*. *J. Bacteriol.* 189:2055-2062.
- Wengelnik, K., Van den Ackerveken, G., and Bonas, U. 1996. HrpG, a key *hrp* regulatory protein of *Xanthomonas campestris* pv. *vesicatoria* is homologous to two-component response regulators. *Mol. Plant-Microbe Interact.* 9:704-712.
- West, A. H., and Stock, A. M. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.* 26:369-376.
- Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene* 87:23-29.