

# *In silico* analysis of the chemotactic system of *Agrobacterium tumefaciens*

Nan Xu<sup>1,2</sup>, Mingqi Wang<sup>1,2</sup>, Xiaojing Yang<sup>1,2</sup>, Yujuan Xu<sup>1,2</sup> and Minliang Guo<sup>1,2,\*</sup>

## Abstract

*Agrobacterium tumefaciens* is an efficient tool for creating transgenic host plants. The first step in the genetic transformation process involves *A. tumefaciens* chemotaxis, which is crucial to the survival of *A. tumefaciens* in changeable, harsh and even contaminated soil environments. However, a systematic study of its chemotactic signalling pathway is still lacking. In this study, the distribution and classification of chemotactic genes in the model *A. tumefaciens* C58 and 21 other strains were annotated. Local BLAST was used for comparative genomics, and HMMER was used for predicting protein domains. Chemotactic phenotypes for knockout mutants of ternary signalling complexes in *A. tumefaciens* C58 were evaluated using a swim agar plate. A major cluster, in which chemotaxis genes were consistently organized as MCP (methyl-accepting chemotaxis protein), CheS, CheY1, CheA, CheR, CheB, CheY2 and CheD, was found in *A. tumefaciens*, but two coupling CheW proteins were located outside the 'che' cluster. In the ternary signalling complexes, the absence of MCP *atu0514* significantly impaired *A. tumefaciens* chemotaxis, and the absence of CheA (*atu0517*) or the deletion of both CheWs abolished chemotaxis. A total of 465 MCPs were found in the 22 strains, and the cytoplasmic domains of these MCPs were composed of 38 heptad repeats. A high homology was observed between the chemotactic systems of the 22 *A. tumefaciens* strains with individual differences in the gene and receptor protein distributions, possibly related to their ecological niches. This preliminary study demonstrates the chemotactic system of *A. tumefaciens*, and provides some reference for *A. tumefaciens* sensing and chemotaxis to exogenous signals.

## DATA SUMMARY

Using the keyword '*Agrobacterium tumefaciens*' in the National Center for Biotechnology Information genome database, *Agrobacterium tumefaciens* and some *Agrobacterium radiobacter* strains (*A. tumefaciens* biovars) were retrieved. Considering the genome sequencing level and related literature report, 21 *A. tumefaciens* strains (C58, P4, 1D1609, cherry 2E-2-2, 186, Ach5, B<sub>6</sub>, CCNWGS0286, DSM30147, F2, GW4, H13, K599, S33, S2, LMG140, LMG125, N273, LBA4404, 5A and WRT31) and *A. radiobacter* K84 were selected. Related genome sequences, 16S rRNA and protein sequences were download from [www.ncbi.nlm.nih.gov/genome](http://www.ncbi.nlm.nih.gov/genome). No new sequence data was generated in this work. An Hidden Markov Model (HMM) model for methyl-accepting chemotaxis proteins (PF00015) and 77 existing ligand-binding domain models (Table S1, available with the online version of this

article) were download from the Pfam database (<http://pfam.xfam.org/>).

## INTRODUCTION

Chemotaxis can help micro-organisms under nutrition stress search for suitable living environments, and is important for host-microbe symbiotic processes, such as host invasion [1], biofilm formation [2] and cell adhesion [3]. The movement of bacteria towards degradable environmental pollutants is related to chemotaxis, and could improve the degradation rates and bioavailability of some pollutants, and effectively promote bioremediation of environmental pollutants [4, 5]. The chemotactic system of bacteria (often called the Che system) is composed of a chemoreceptor and core proteins. Genes encoding core proteins always occur in clusters in bacterial genomes, and mostly in one operon.

Received 10 July 2020; Accepted 06 October 2020; Published 29 October 2020

**Author affiliations:** <sup>1</sup>College of Bioscience and Biotechnology, Yangzhou University, Yangzhou 225009, PR China; <sup>2</sup>Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of Education of China, Yangzhou University, Yangzhou 225009, PR China.

\*Correspondence: Minliang Guo, [guoml@yzu.edu.cn](mailto:guoml@yzu.edu.cn)

**Keywords:** *Agrobacterium tumefaciens*; chemotaxis; Che cluster; methyl-accepting chemotaxis protein.

**Abbreviations:** LBD, ligand-binding domain; MCP, methyl-accepting chemotaxis protein; NCBI, National Center for Biotechnology Information; SD, signalling domain; TM, transmembrane helix.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Three supplementary tables are available with the online version of this article.

000460 © 2020 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

Chemoreceptors, also known as methyl-accepting chemotaxis proteins (MCPs), are front-end signal acceptors, which specifically sense chemokine effectors and transmit sensed signals to downstream proteins of the Che system [6]. In most bacteria, core proteins are highly conserved, but different MCPs could recognize specific chemoattractants. The most extensively studied Che system is the *Escherichia coli* system [7, 8], composed of five MCPs (Tar, Tsr, Trg, Aer and Tap) and six core proteins (CheA, CheB, CheR, CheW, CheY and CheZ).

*Agrobacterium tumefaciens*, a Gram-negative alphaproteobacterium, is abundant in soil. *Agrobacterium* has a complicated taxonomic classification [9], which was derived from the evolutionary relationships between the 16S rRNA marker, *recA* genes and other conserved housekeeping genes using important public databases such as the National Center for Biotechnology Information (NCBI), SILVA [10], RDP [11] and IMG [12] databases. The well-accepted classification includes three biovars: *A. tumefaciens* species complex (biovar I), *Agrobacterium rhizogenes* (biovar II) and *Agrobacterium vitis* (biovar III) [13]. As a common plant pathogen, *A. tumefaciens* was initially identified as the cause of plant tumours. In *A. tumefaciens*, Braun first proposed, in 1947, the tumour-inducing principle that meant that some of the bacterial components might be tumorigenic factors [14]. DNA fragments of a tumour-inducing (Ti) plasmid were later validated to be the tumour-inducing principle delivered by *A. tumefaciens* into plant cells [15]. T-DNA and most virulence genes are located on the Ti plasmid. *A. tumefaciens* is capable of transferring part of its virulent Ti plasmid into plant cells and integrating the segments into the host genome, thereby implementing genetic transformation of host cells. To increase the T-DNA transfer frequency of *A. tumefaciens*, various *A. tumefaciens* mutants were constructed to increase *vir* gene expression or remove adverse factors from *Agrobacterium*–plant interactions. For example, super-*Agrobacterium* ver. 4, containing the ACC deaminase (*acdS*) and GABA transaminase (*gabT*) genes, was validated as a more effective and powerful tool for plant genetic engineering than the previously developed strains [16].

T-DNA transfer mediated by *A. tumefaciens* has been widely used in plant transgenic technology [17], and the microbe's chemotaxis is directly associated with the onset of its transgenic mechanism. The first step in the *A. tumefaciens* infection process is to recognize suitable infection sites in plant hosts by chemotactic activity [18], and previous studies of chemotaxis in *A. tumefaciens* have focused on the identification of attractant types, rarely investigating the chemotaxis mechanism. Some organic compounds secreted by plant callus could be attractants of *A. tumefaciens* and induce the expression of oncogenes [19, 20]. *A. tumefaciens* is reported to be sensitive to many sugars, phenols and amino acids [21, 22], some of which strongly induce chemotaxis and tumorigenesis [23–25]. However, these studies are scattered and lack systematic research [26, 27].

### Impact Statement

Chemotaxis is the phenomenon by which motile microorganisms sense the presence and/or concentration of chemicals and move towards favourable attractants or avoid harmful repellents. The first step in the *Agrobacterium tumefaciens* infection and transgenic process is driven by chemotactic activity. By comparative genomics analysis, one chemotactic operon was found to exist in *A. tumefaciens*, and the order of chemotactic genes was generally MCP (methyl-accepting chemotaxis protein), CheS, CheY1, CheA, CheR, CheB, CheY2 and CheD, except for GW4 and B<sub>6</sub> strains. Two CheW coupling proteins in *A. tumefaciens* were outside the chemotactic operon. Formation of the ternary signalling complex (MCP-CheA-CheW) was the basis for chemotaxis in *A. tumefaciens* C58. In both CheA single mutation strains and CheWs double deficient strains chemotaxis was eliminated. The number of MCPs in most *A. tumefaciens* strains exceeded 20, and they belonged to classes I–IV, excluding class II. Class I had the largest number of MCPs, and the Ia subtype had two transmembrane domains, whereas the Ib subtype had one transmembrane domain. Among class IV cytoplasmic MCPs, the IVa subtype had more transmembrane domains than the IVb subtype. About 80% of *A. tumefaciens* MCPs had ligand-binding domains (LBDs), belonging to 15 different domain models, classified into four domain superfamilies: 4HB, Cache, PAS and protoglobin. Cache, the most abundant LBD superfamily in *A. tumefaciens*, was found to bind to amine ligands. The 4HB superfamily covered all the domain models, namely 4HB\_MCP\_1, CHASE3, TarH and HBM. PAS and protoglobin superfamilies were only found in cytoplasmic MCPs, which might be related to intracellular redox state induction. Our study explains the chemotaxis system of *A. tumefaciens*, and serves as a basis for future studies on *A. tumefaciens* sensing and chemotaxis to exogenous signals.

Chemotaxis not only affects survival under stress and in polluted environments, but also is closely related to host infection. *A. tumefaciens* possesses two contrasting lifestyles: independent saprophytic or pathogenic. Considering the importance of chemotaxis to *A. tumefaciens*, it was expected to become a model bacterium in chemotaxis research. In this study, the Che system and MCPs of the model strain, *A. tumefaciens* C58, were annotated. The effect of the ternary signalling complex on chemotaxis was validated, as an example of the MCP located on the chemotactic operon. Furthermore, the chemotactic operons of other *A. tumefaciens* strains were compared, and their MCP types and the diversity of the ligand-binding domain (LBD) were systematically investigated.

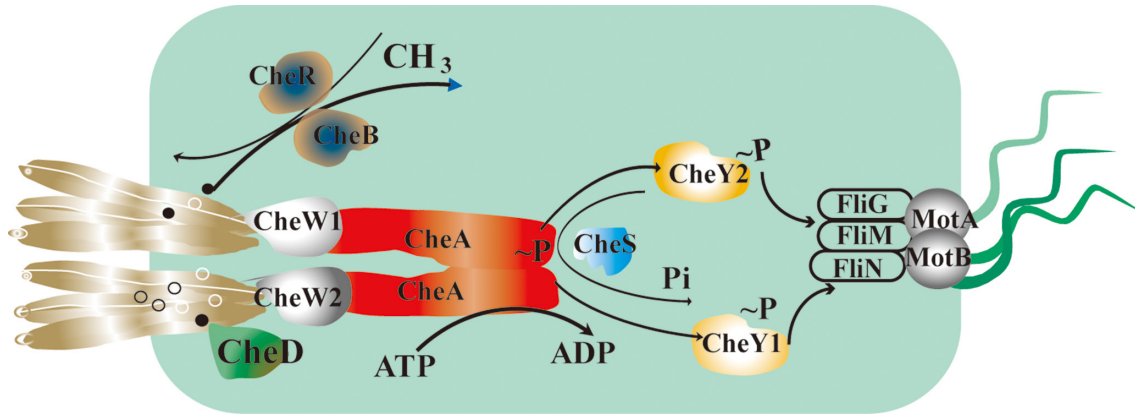


Fig. 1. Chemotactic pathway in *A. tumefaciens* C58.

## METHODS

### Multiple sequence alignment and phylogenetic analyses

NCBI nucleotide BLAST was used for homologous comparison against the local library of *che* genes in *A. tumefaciens* C58. Default parameter values were >90% similarity and  $<10^{-30}$  *E* value. MEGA X software was used to reconstruct 16S rRNA gene phylogenetic trees by the neighbour-joining method [28].

### Protein analysis of MCPs

HMMER [29] was used to interrogate all protein sequences of the *A. tumefaciens* strains for matches to Pfam PF00015. The proteins with this domain were identified as MCPs. Protein domains of MCPs were preliminarily predicted by the SMART server (<http://smart.embl.de/>). Transmembrane domains of MCPs were determined by the TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM/>). LBDs of MCPs were determined by comparing 77 existing LBDs via HMMER.

### Construction of *A. tumefaciens* deletion mutants and complementation mutants

Using diluted bacterial liquid as the template, the upstream sequence of target genes was amplified by using DgeneU-F and DgeneU-R primers, and the downstream sequence of target genes was amplified by using DgeneD-F and DgeneD-R primers (see Table S2). Here, the gene refers to 0514, 0517, W1 and W2 in Table S2. The target fragment was obtained by overlap PCR using DgeneU-F and DgeneD-R primers. The target fragments were identified via agarose gel electrophoresis (1%) and were recovered using a TaKaRa agarose gel DNA extraction kit. The purified PCR products and suicide vector pEX18Km were digested, linked and ligated, and used to transform *E. coli* DH5 $\alpha$ . After PCR and DNA sequencing, the correctly constructed plasmids were used to transform *A. tumefaciens* C58. Deletion mutants were selected and validated using a kanamycin-resistance gene

as the positive selection marker and a suicide gene, *sacB*, as the counter-selectable marker.

Using diluted bacterial liquid as the template, target genes with their native promoters (upstream 500 bp) were amplified using Hgene-F and Hgene-R primers (see Table S2). The identified and purified PCR product was inserted into plasmid pCB301. Correctly constructed plasmids were used to transform the corresponding deletion mutation strains of *A. tumefaciens*. The success of the complementation was verified by detecting kanamycin resistance and PCR sequencing. All the strains, plasmids, culture media and primers used in this study are listed in Tables S1–S3.

### Chemotaxis assays using swim agar plates

Chemotactic responses to nutrient substances were assayed by performing an agar swim test [30]. The biomass in the mid-exponential growth phase was adjusted to OD<sub>600</sub> 0.5. Cell cultures (3  $\mu$ l) were added to a swim agar plate (0.18% agar) containing AB-sucrose medium. After incubation at 28 °C for 48 h, the colony size was observed and recorded in photos.

## RESULTS AND DISCUSSION

### Annotation of the chemotaxis system of *A. tumefaciens* C58

Bacterial chemotaxis generally involves histidine kinase CheA, coupling protein CheW, response regulator CheY, MCP and flagellar motor proteins. There was only one distinct chemotaxis operon located on the *A. tumefaciens* C58 circular chromosome (from *atu0514* to *atu0521*), which contained some two-component Che proteins. However, coupling protein CheW was not in the operon, but in other positions on the circular chromosome (CheW1 *atu2075* and CheW2 *atu2617*). There were two homologous methyltransferase CheR proteins; one (*atu0518*) was found in the operon and the other (*atu4805*) was on the linear chromosome. In addition, 20 genes were predicted to encode MCPs in the C58 genome;

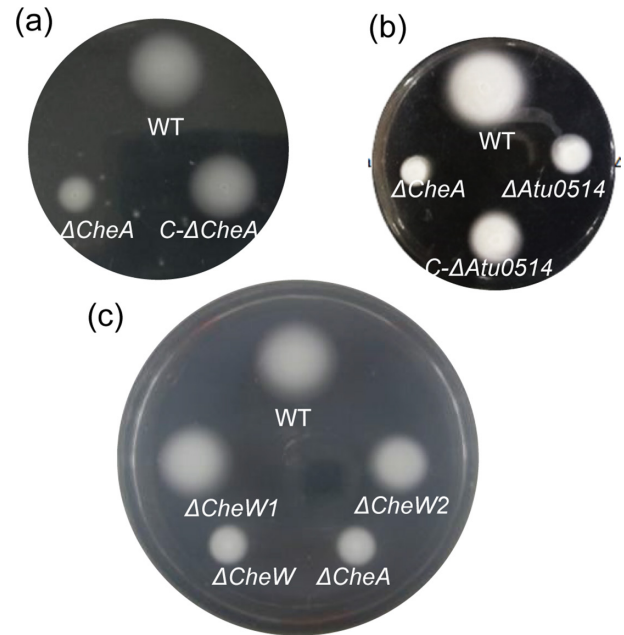


atu0514 was found in the chemotaxis operon, whereas the others were scattered in the Ti plasmid (atu6132), At plasmid (atu5442), circular chromosome (atu0373, atu0387, atu0526, atu0646, atu0738, atu0872, atu1027, atu1912, atu2173, atu2223, atu2360, atu2618), and linear chromosome (atu3094, atu3330, atu3363, atu3725, atu4736).

CheA (atu0517) was the core component of the chemotaxis operon in *A. tumefaciens* C58. Among three histidine kinase categories, atu0517 belonged to group I, including the histidine-containing phosphotransferase (HPt) domain, homodimeric domain of histidine kinase (H-kinase\_dim, 337–400), histidine kinase-like ATPase domain (HATPase\_c, 445–586) and CheA regulatory domain (CheA\_reg, 578–719). The chemotaxis signal pathway of *A. tumefaciens* C58 is shown in Fig. 1. CheA activity can be regulated by methylation and demethylation of methyltransferase CheR (atu0518) and methylesterase CheB (atu0519), respectively, and activated by removing specific glutamine residues from MCPs of glutamine deamidase CheD (atu0521). CheA phosphorylates CheY1 (atu0516) and CheY2 (atu0520). Phosphorylated CheY interacts with the flagellum motor complex (Flis/Mots) and drives flagellum rotation in response to chemoattractants. Like for other *Alphaproteobacteria*, phosphatase CheZ was not found in the *A. tumefaciens* C58 genome [31]. The second gene, atu0515, in the chemotaxis operon was annotated as CheS, which works with CheA to dephosphorylate CheY.

### Chemotaxis of mutants of the ternary signalling complex in *A. tumefaciens* C58

The colony size on the swim agar plate is often used to examine chemotaxis behaviour [30]. When bacterial cells are inoculated onto swim agar, they utilize the nutrients around the inoculation area. If bacterial cells show a sensitive chemotactic response, they move outward along the nutrient gradient and form large colonies. If their chemotactic system is damaged, the colonies are small. CheW1 (atu2075) and CheW2 (atu2617) couple CheA dimer to two chemoreceptor trimers of dimers to form a chemoreceptor-CheW-CheA ternary signalling complex. External chemical substances and signal stimuli can be sensed by the ternary signalling complex allowing movement towards favourable environmental conditions and avoidance of harmful chemoattractant repellents. The homologous recombination method with a suicide gene, *sacB*, as the counter-selectable marker was used to delete an MCP (atu0514), CheW1 (atu2075), CheW2 (atu2617) and CheA (atu0517). Mutant colony sizes were detected on a swim agar plate. As shown in Fig. 2, chemotaxis ability was lost after deleting CheA but recovered to the level of the wild-type by the introduction of an expression plasmid, pCB301, harbouring complete CheA. The colony size of  $\Delta$ atu0514 was significantly smaller than that of the wild-type, and larger than that of  $\Delta$ CheA, suggesting that knockout of atu0514 would weaken *A. tumefaciens* C58 chemotaxis. Colonies of C58 $\Delta$ w1 and C58 $\Delta$ w2 mutants were smaller than those of wild-type C58, but larger than those of  $\Delta$ CheA. The colony size of the double-mutant ( $\Delta$ CheW) was almost the same as that of  $\Delta$ CheA. One CheW could partially compensate the

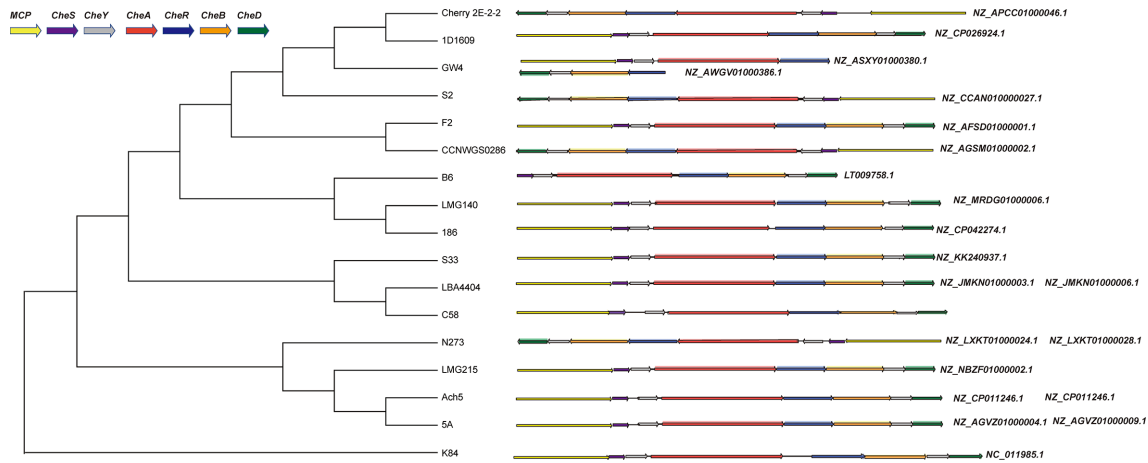


**Fig. 2.** Swim plate colonies of *A. tumefaciens* C58 strains with mutations of the ternary signalling complex.

effects of the other CheW on chemotaxis, but in knockouts of both, CheWs chemotaxis was completely lost. In that case, CheA were not bounded to cell poles by CheWs but dispersed in cells, so that the ternary signalling complex could not form, and chemotaxis signal transduction was blocked.

### Characteristics of MCPs in *A. tumefaciens* C58

The 20 MCPs predicted in *A. tumefaciens* C58 were a far higher number than those found in the model microorganism, *E. coli* (5 MCPs), and its close relative *Sinorhizobium meliloti* [32]. The large number of MCPs indicates that *A. tumefaciens* could recognize more chemicals for survival in complex environments. According to the number of heptad repeats, chemoreceptors were classified into seven categories (24, 28, 34, 36, 38, 40, and 44h) [33]. All the MCPs in *A. tumefaciens* C58 belonged to the 36h type. Six MCPs without transmembrane regions accounted for 30% of all the MCPs in *A. tumefaciens*, and the ratio was between the mean value of MCPs in bacteria (~14%) and archaea (~43%). Some domains in the six soluble chemoreceptor proteins could sense small molecules or redox-sensitive cofactors. For example, two PAS domains, the most abundant domain in soluble MCPs, were located at the N-terminal of the four MCPs [34]. Proto-globin was located at the N-terminal of the other two MCPs, belonging to the haemoglobin family, which could bind with the prosthetic group of haem [35]. Among 14 transmembrane MCPs, atu0387 and atu3330 had only one transmembrane region, 7 MCPs had two transmembrane regions, and 5 MCPs had three transmembrane regions. Nine of the fourteen MCPs were annotated with the periplasmic LBD, which were distributed in three types. Double Cache (atu1912, atu2173, atu0526,



**Fig. 3.** Organization of the chemotaxis operons in *A. tumefaciens* strains and a phylogenetic tree reconstructed based on published 16S rRNA sequences of *A. tumefaciens*.

atu2223, atu0373 and atu3725) consisted of two  $\alpha/\beta$  subdomains and a long N-terminal helix. It is the most common extracellular domain in prokaryotes [36]. The single Cache domain of atu0646 was similar to the double Cache domain, both of which belong to the Cache-like superfamily. A single 4HB domain existed in atu0738 and atu0387, comprising two symmetric antiparallel coiled coils.

### Comparison of the chemotaxis operon in *A. tumefaciens* strains

Among the 22 *A. tumefaciens* strains, the 16S rRNA sequences of 17 were publicly available in the NCBI database; phylogenetic trees were reconstructed based on these 16S rRNA sequences. As shown in Fig. 3, 16 different strains of *A. tumefaciens* were clustered into one branch, and *Agrobacterium radiobacter* K84 was on a separate branch. Comparing the *che* gene cluster of the 17 strains, the distribution of chemotaxis genes appears consistent with that of *A. tumefaciens* C58 in this order: CheS, CheY1, CheA, CheR, CheB, CheY2 and CheD. However, the chemotaxis genes of *A. tumefaciens* GW4 were distributed on different scaffolds, i.e. MCP, CheS, CheY1, CheA and CheR on NZ\_AWGV01000458.1; CheR, CheB, CheY2 and CheD on NZ\_AWGV01000386.1 in reverse arrangement. Without considering the quality of published genome sequences, this distribution might be related to GW4 being isolated from As-contaminated aquifer sediments [37]. No MCP like atu0514 was found in the chemotaxis operon of non-pathogenic *A. tumefaciens* B<sub>6</sub> [38]. All the 17 strains had two CheWs outside the chemotaxis operon. In addition, CheR had other homologous genes outside the chemotaxis operons in the C58, LBA4401, N2/73, LMG215, Ach5 and 5A strains.







### Distribution of MCPs in *A. tumefaciens* strains

MCPs typically include a cytoplasmic signalling domain (SD), transmembrane helices (TMs) and a LBD [39]. The SD, the most conservative element, consists of a methylation helices subdomain (MH), signalling subdomain (SSD) and flexible

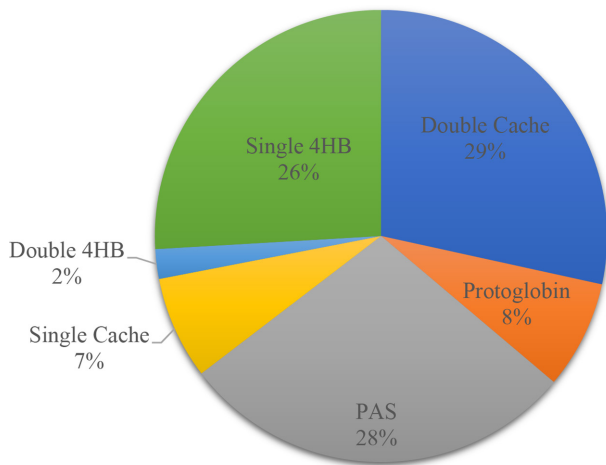
bundles subdomain (FB) between both subdomains. Sites for the interaction of MCPs and Che proteins (CheA, CheW, CheR, CheB and CheD) are located in the SD region. The LBD widely exists in the cytoplasm or periplasm, and recognizes extracellular or intracellular chemical and physical signals. Chemoreceptors for extracellular signals have at least one TM and one periplasmic LBD. Based on the membrane topology, MCPs are divided into four major classes (I–IV) and seven subclasses (Ia, Ib, II, III<sub>m</sub>, III<sub>c</sub>, IV<sub>a</sub>, IV<sub>b</sub>) [40].

A total of 465 MCPs were predicted in the 22 isolates of *A. tumefaciens*. Most strains had 20–25 MCPs, there were 22 MCPs in five strains, 21 MCPs in four strains, and 20 MCPs in four strains (Table 1). *A. tumefaciens* S2 and *A. radiobacter* K84 had only 18 MCPs, while *A. tumefaciens* DSM 30147 had up to 28 MCPs, possibly related to its isolation from saprobic soil, and need to adapt to a more complex and changeable environment. The 465 MCPs belonged to three classes, excluding class II, which has a cytoplasmic LBD at the N-terminal connected with a cytoplasmic SD via two TMs. However, class II receptors only account for about 3% of present MCPs, far less than other classes. The predominant MCPs of *A. tumefaciens* were the class I MCPs (200). Typical structures in class I included TMs, a periplasmic LBD located between TMs and a cytoplasmic SD, which were further categorized into two subclasses: Ia (two TMs) and Ib (one TM). Ia subclass MCPs were more common than Ib MCPs. A total of 198 subclass Ia and 2 subclass Ib MCPs were found in *A. tumefaciens*. The second most common MCPs were the class III MCPs with 1–8 TMs. Among the 144 class III MCPs of *A. tumefaciens*, 103 MCPs with sensor elements were located within the membrane (III<sub>m</sub> subclass), and the cytosolic LBDs of 41 MCPs were located after the last TMs region (III<sub>c</sub> subclass). A total of 121 MCPs belonged to class IV, which were soluble cytoplasmic MCPs without TMs. A large majority belonged to the IV<sub>a</sub> subclass with LBD regions, and four MCPs without LBD regions belonged to IV<sub>b</sub>.

**Table 1.** Classification of MCPs from *A. tumefaciens* strains

	186	ID1609	2E-2-2	H13	K599	K84	5A	Ach5	LBA404	LMG125	LMG140	B6	C58	N273	P4	S2	S33	CCNWGS0286	DSM30147	F2	GW4	WRT31	
Ia 	8(I) 2(II)	4(I) 4(II)	6(I) 2(II)	7(I) 3(II)	7(I)	7(I)	7(I) 1(II)	9(I) 3(II)	7(I) 2(II)	6(I) 2(II)	8(I) 2(II)	8(I) 2(II)	5(I) 1(II)	9(I) 3(II)	7(I) 3(II)	9(I) 2(II)	9(I) 2(II)	7(I) 2(II)	9(I)	10(I) 1(II)	5(I) 3(II)	4(I) 3(II)	6(I) 3(II)
Ib 									1									1					
IIIm 	4	5	3	3	3	5	5	5	4	5	5	6	7	3	4	2	5	4	9	3	7	6	6
IIIfc 	2	1	2	3	5	4	1	1	2	2	2	1	1	2	2	1	1	2	3	2	2	1	1
Iva 	6	6	6	5	4	2	7	6	5	6	5	5	6	5	6	4	5	5	5	6	5	7	7
Ivb 				1	1	1	1	1			1												
Total	22	20	19	22	20	18	20	25	21	21	22	23	20	22	22	18	19	21	28	19	21	21	23

Based on ligand-binding domain (LBD) length, the Ia subclass can be divided into type I (170-215 aa) and type II (215-299 aa); the first line is the strain number (the names are omitted, see the material method for details).



**Fig. 4.** Proportions of the six types of LBD in *A. tumefaciens*.

The different subtypes of MCPs in the 22 isolates of *A. tumefaciens* are listed in Table 1. Types Ia(I), III(m) and IV(a) were found in all 22 strains, followed by III(c) and Ib(I) in 20 and 19 strains, respectively; the least frequent were IV(b) and Ib, which were only found in 4 and 2 strains, respectively. Based on LBD length, the Ia subclass can be divided into type I (120–215 aa) and type II (215–299 aa).

### Analysis of LBDs in *A. tumefaciens* strains

Further analyses of the 465 MCPs in the 22 *A. tumefaciens* showed that 100 MCPs did not have LBD regions or could not be matched with existing LBDs in the Pfam database. At least one region on the remaining 365 MCPs was an identifiable LBD, which covered 516 LBDs, but was not evenly distributed. More than 95% of MCPs had only one LBD, and about 2% of MCPs had two LBD segments. The highest number contained was five LBDs, such as CDN92820.1 of *A. tumefaciens* S2 (PAS\_3 34–119, 155–240, and CHASE3 269–300, 353–391, 500–551) and TQO46244.1 of *A. tumefaciens* 5A (PAS\_3 34–119, 156–240, and CHASE3 269–360, 332–392, 498–558). There were 15 different LBD models in the *A. tumefaciens* MCPs belonging to six types: single 4HB, double Cache, double 4HB, single Cache, PAS and protoglobin. As shown in Fig. 4, there were four domain superfamilies: Cache (36%), 4HB (28%), PAS (28%) and protoglobin (8%). Like other micro-organisms, the three most abundant superfamilies (4HB, Cache and PAS) accounted for more than 80% of known LBDs in *A. tumefaciens*. The Cache superfamily was the most abundant LBD superfamily; there was a higher proportion of double Cache than single Cache. Double Cache, with a bimodular arrangement, was composed of two PAS-like modules and a long N-terminal  $\alpha$ -helix, whereas single Cache had only one PAS-like module. Double Cache has been reported to bind to the membrane-distal module rather than the membrane-proximal module [41], and select amines as effectors [42]. Atu0526 (Cache\_3–Cache\_2: 78–197) and atu1912 (double Cache\_2:32–182) were validated to be MCPs for some amino acids in *A. tumefaciens* C58 (data not shown).

*A. tumefaciens* had all the domain models of the 4HB superfamily. The single 4HB type included three domain models: 4HB\_MCP\_1 (55), CHASE3 (73) and TarH (6). There were 11 HBM domain models in the double 4HB type. Many kinds of 4HB superfamily ligands have been found, among which the HBM domain mainly recognizes intermediates of the TCA cycle [43]. The PAS superfamily was also a common LBD in the bacterial signal transduction system. Unlike the Cache and 4HB families, the PAS of *A. tumefaciens* only exists in cytosolic MCPs. The binding sites of haem, FAD or FMN were found in the PAS domain, which might be a function of aerotaxis and redox sensing [44].

### Funding information

This work was supported by the National Natural Science Foundation of China (21808196, 31870118), the China Postdoctoral Science Foundation (2018M632389) and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (18KJB180030).

### Author contributions

N. X. and M. G. conceived the project. N. X. and M. W. completed data collection and data analysis. N. X. and Y. X. completed wet experiments. N. X., M. W., X. Y. and M. G. wrote the manuscript.

### Conflicts of interest

The authors declare that there are no conflicts of interest.

### References

- Erhardt M. Strategies to block bacterial pathogenesis by interference with motility and chemotaxis. *Curr Top Microbiol Immunol* 2016;398:185–205.
- Alexandre G. Chemotaxis control of transient cell aggregation. *J Bacteriol* 2015;197:3230–3237.
- Leonard S, Hommais F, Nasser W, Reverchon S. Plant-phytopathogen interactions: bacterial responses to environmental and plant stimuli. *Environ Microbiol* 2017;19:1689–1716.
- Mangwani N, Kumari S, Das S. Bacterial biofilms and quorum sensing: fidelity in bioremediation technology. *Biotechnol Genet Eng Rev* 2016;32:43–73.
- Adadevoh JS, Triolo S, Ramsburg CA, Ford RM. Chemotaxis increases the residence time of bacteria in granular media containing distributed contaminant sources. *Environ Sci Technol* 2016;50:181–187.
- Jones CW, Armitage JP. Positioning of bacterial chemoreceptors. *Trends Microbiol* 2015;23:247–256.
- Hazelbauer GL. Bacterial chemotaxis: the early years of molecular studies. *Annu Rev Microbiol* 2012;66:285–303.
- Parkinson JS, Hazelbauer GL, Falke JJ. Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. *Trends Microbiol* 2015;23:257–266.
- Gan HM, Savka MA. One more decade of *Agrobacterium* taxonomy. *Curr Top Microbiol Immunol* 2018;418:1–14.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–D596.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014;42:D633–D642.
- Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 2012;40:D115–D122.
- Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK et al. Genome sequences of three *agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J Bacteriol* 2009;191:2501–2511.



14. Braun AC. A physiological basis for autonomous growth of the crown-gall tumor cell. *Proc Natl Acad Sci USA* 1958;44:344–349.
15. Chilton MD, Montoya AL, Merlo DJ, Drummond MH, Nutter R et al. Restriction endonuclease mapping of a plasmid that confers oncogenicity upon *Agrobacterium tumefaciens* strain B6-806. *Plasmid* 1978;1:254–269.
16. Nonaka S, Someya T, Kadota Y, Nakamura K, Ezura H. Super-*Agrobacterium* ver. 4: improving the transformation frequencies and genetic engineering possibilities for crop plants. *Front Plant Sci* 2019;10:1204.
17. Lacroix B, Citovsky V. Transfer of DNA from bacteria to eukaryotes. *mBio* 2016;7:e00863-16.
18. Hawes MC, Smith LY. Requirement for chemotaxis in pathogenicity of *Agrobacterium tumefaciens* on roots of soil-grown pea plants. *J Bacteriol* 1989;171:5668–5671.
19. Winans SC. Two-way chemical signaling in *Agrobacterium*-plant interactions. *Microbiol Rev* 1992;56:12–31.
20. Shaw CH. *Agrobacterium tumefaciens* chemotaxis protocols. *Methods Mol Biol* 1995;44:29–36.
21. Ashby AM, Watson MD, Loake GJ, Shaw CH. Ti plasmid-specified chemotaxis of *Agrobacterium tumefaciens* C58C1 toward vir-inducing phenolic compounds and soluble factors from monocotyledonous and dicotyledonous plants. *J Bacteriol* 1988;170:4181–4187.
22. Loake GJ, Ashby AM, Shaw CH. Attraction of *Agrobacterium tumefaciens* C58C1 towards sugars involves a highly sensitive chemotaxis system. *Microbiology* 1988;134:1427–1432.
23. Bolton GW, Nester EW, Gordon MP. Plant phenolic compounds induce expression of the *Agrobacterium tumefaciens* loci needed for virulence. *Science* 1986;232:983–985.
24. Cangelosi GA, Ankenbauer RG, Nester EW. Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. *Proc Natl Acad Sci USA* 1990;87:6708–6712.
25. Stachel SE, Messens E, Van Montagu M, Zambryski P. Identification of the signal molecules produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. *Nature* 1985;318:624–629.
26. Harighi B. Genetic evidence for CheB- and CheR-dependent chemotaxis system in *A. tumefaciens* toward acetosyringone. *Microbiol Res* 2009;164:634–641.
27. Merritt PM, Danhorn T, Fuqua C. Motility and chemotaxis in *Agrobacterium tumefaciens* surface attachment and biofilm formation. *J Bacteriol* 2007;189:8005–8014.
28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
29. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R et al. HMMER web server: 2018 update. *Nucleic Acids Res* 2018;46:W200–W204.
30. Sampedro I, Parales RE, Krell T, Hill JE. *Pseudomonas* chemotaxis. *FEMS Microbiol Rev* 2015;39:17–46.
31. Guo M, Huang Z, Yang J. Is there any crosstalk between the chemotaxis and virulence induction signaling in *Agrobacterium tumefaciens*? *Biotechnol Adv* 2017;35:505–511.
32. Miller LD, Yost CK, Hynes MF, Alexandre G. The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *viciae* is essential for competitive nodulation. *Mol Microbiol* 2007;63:348–362.
33. Alexander RP, Zhulin IB. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proc Natl Acad Sci USA* 2007;104:2885–2890.
34. Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* 1999;63:479–506.
35. Hou S, Larsen RW, Boudko D, Riley CW, Karatan E et al. Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature* 2000;403:540–544.
36. Upadhyay AA, Fleetwood AD, Adebali O, Finn RD, Zhulin IB. Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. *PLoS Comput Biol* 2016;12:e1004862.
37. Fan H, Su C, Wang Y, Yao J, Zhao K et al. Sedimentary arsenite-oxidizing and arsenate-reducing bacteria associated with high arsenic groundwater from Shanyin, Northwestern China. *J Appl Microbiol* 2008;105:529–539.
38. Rapp BJ, Kemp JD, White F. Isolation of a non-tumor-inducing mutant of the Ti plasmid of *Agrobacterium tumefaciens* strain B<sub>6</sub>. *Can J Microbiol* 1979;25:291–297.
39. Bi S, Lai L. Bacterial chemoreceptors and chemoeffectors. *Cell Mol Life Sci* 2015;72:691–708.
40. Salah Ud-Din AIM, Roujeinikova A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cell Mol Life Sci* 2017;74:3293–3303.
41. Nishiyama SI, Takahashi Y, Yamamoto K, Suzuki D, Itoh Y et al. Identification of a *Vibrio cholerae* chemoreceptor that senses taurine and amino acids as attractants. *Sci Rep* 2016;6:20866.
42. Corral-Lugo A, De la Torre J, Matilla MA, Fernández M, Morel B et al. Assessment of the contribution of chemoreceptor-based signalling to biofilm formation. *Environ Microbiol* 2016;18:3355–3372.
43. Lacial J, Alfonso C, Liu X, Parales RE, Morel B et al. Identification of a chemoreceptor for tricarboxylic acid cycle intermediates: differential chemotactic response towards receptor ligands. *J Biol Chem* 2010;285:23126–23136.
44. Xie Z, Ulrich LE, Zhulin IB, Alexandre G. PAS domain containing chemoreceptor couples dynamic changes in metabolism with chemotaxis. *Proc Natl Acad Sci USA* 2010;107:2235–2240.

### Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at [microbiologyresearch.org](http://microbiologyresearch.org).