Supporting Information

Briegel et al. 10.1073/pnas.0905181106

SI Text

Culture Conditions. *C. crescentus* CB15N cells were grown at 30 °C in peptone yeast extract medium (1) to log phase. Two milliliters of the cell suspension was pelleted for 5 min at a relative centrifugal force (rcf) of 1,500. The pellet was rinsed in M2 salt solution (1) and then re-suspended in 30 to 50 μL of fresh M2 solution.

E. coli (RP437 and MG1655) was grown to saturation overnight at 30 °C in T-broth (10 g bacto tryptone and 5 g NaCl in 1 L); 0.1 mL of this culture was used to inoculate 10 mL S medium (T-broth, 0.1% $MgSO_4$, 12% sucrose, 0.1 mg thiamine/100 mL, sterilized by filtration) and grown for 4.5 h at 30 °C to log phase; and 3.3 mL of this culture was added to 6.6 mL S-medium containing 700 IU/mL penicillin G potassium salt and incubated for approximately 60 min at 30 °C (2).

T. maritima (MSB8; DSM 3109) was grown at 80 °C overnight under strictly anaerobic conditions in a slightly modified Murashige and Skoog medium as described (3). After cooling the cell culture to room temperature, several milliliters were spun down at 4 °C for 5 min at 18,000 × *g* and resuspended in a few hundred microliters of supernatant.

For polymyxin B treatment, 10 mL of cell culture were spun down at 4 °C for 5 min at a speed of 18,000 × *g*. The pellet was resuspended in 2 mL fresh medium on ice. After adding 2 mg polymyxin B, the cells were incubated at 4 °C for 10 h before being plunge-frozen onto EM grids.

V. cholerae TRH7000 cells, containing a pMMB67 plasmid overexpressing MreB-GFP or GFP-MreB, were grown in M9 growth medium [M9 minimal salts (Difco), 0.4% glucose, 0.4% casamino acids, 1.0 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 mg/mL thymine] at 37 °C until they reached log phase (OD_{600} = 0.3). Both induced and un-induced cells were imaged and analyzed.

M. magneticum sp. AMB-1 cultures were grown in MG medium as described previously (4).

H. hepaticus (ATCC 51449) was grown on trypticase soy blood agar plates for 3 d, with all 3 days or just the final day being in the presence of added gut epithelial host cells. Cells were scraped off the plate and centrifuged at 7,000 rcf for 5 to 10 min to separate host cells from bacteria.

C. jejuni (ATCC 29428) was initially grown on plates with *Brucella* growth medium supplemented with 5% sheep blood at 37 °C, an O_2 concentration of 1.5%, and a CO_2 concentration of 10%. A colony was then transferred into liquid medium and grown overnight in brain heart infusion broth plus 5% fetal bovine serum at 37 °C and 1.5% O_2.

R. sphaeroides (NCIB 8253) was grown aerobically at 30 °C overnight in LB media.

Borrelia burgdorferi B31 cells (ATCC 35210) were grown to late log phase in Barbour-Stoenner-Kelly II medium with 6% rabbit serum at 35 °C (5). One milliliter of the cell suspension was pelleted for 5 min at 1,500 rcf and resuspended in 400 μL supernatant.

Listeria monocytogenes strain 10403S (serotype 1/2a) that has defined, in-frame deletions in the actA and plcB genes for safety purposes was grown in BHI broth overnight (6). One milliliter cell culture was pelleted for 5 min at 1,500 rcf and resuspended in approximately 50 μL supernatant.

A. longum APO-1 (DSM 6540) and *T. primitia* strain ZAS-2 were grown at 23 °C in sealed culture tubes containing 4YACo medium under an atmosphere of 80% H_2 and 20% CO_2 as described previously (7). Cells were harvested during log-growth at an OD_{600} of approximately 0.6.

H. neapolitanus C2 (ATCC 23641) cells were grown in flasks or bio-reactors in the laboratory of Gordon Cannon (University of Southern Mississippi, Hattiesburg, MS) under previously described conditions (8, 9).

EM Sample Preparation, Data Collection, and Image Processing. EM R2/2 copper/rhodium Quantifoil or lacy carbon grids were glow-discharged and coated with a 3×-concentrated solution of 10-nm colloidal gold particles (Ted Pella). A 5×-concentrated solution of 10-nm colloidal gold was added to the cells immediately before plunge freezing. A 4-μL droplet of the sample solution was applied to the EM grid, then automatically blotted and plunge-frozen into liquid ethane (10) or into a liquid ethane-propane mixture (11) using a Vitrobot (FEI Company) (12) or in-house plunger. The grids were stored under liquid nitrogen until data collection.

EM images were collected using a Polara 300-kV FEG transmission electron microscope (FEI Company) equipped with an energy filter (slit width 20 eV; Gatan) on a 2k × 2k Ultrascan CCD camera or, later, a lens-coupled 4k × 4k UltraCam (Gatan). Pixels on the CCD represented between 0.64 nm and 1.26 nm on the specimen. Typically, tilt series were recorded from -60° to 60° with an increment of 1° semi-automatically around 1 or 2 axes (13) at 10 and 12 μm under-focus using the predictive UCSF-Tomo package (14) or Leginon (15). A cumulative dose as high as 200 $e^-/Å^2$ was used. The data from *T. maritima* were collected in the laboratory of Wolfgang Baumeister (Max Planck Institute for Biochemistry, Martinsried, Germany) using a Philips CM 300 transmission electron microscope equipped with a field emission gun and a GIF 2002 energy filter (Gatan).

Three-dimensional reconstructions were calculated using the IMOD software package (16). The *T. maritima* reconstructions were generated by weighted back-projection with the EM software package (17). A median filter was applied to noisier data sets to enhance the signal-to-noise ratio (16).

MCP Classification. MCPs were assigned to signaling classes and membrane topology types as previously described (18, 19). Briefly, each sequence was matched to HMM models of the established MCP signaling classes using HMMER. Sequences that did not match any established signaling class were left unclassified ("unc" in Table 1). To determine the MCP membrane topology type, each sequence was scanned with the SignalP (20) and DAS (21) programs to reveal the signal peptide and transmembrane regions (TMs), respectively. The Phobius package (22) was used to independently verify TM predictions with the TMHMM2 algorithm. The often under-predicted second TM (TM2) in MCP topology type I receptors was further confirmed by aligning HAMP domain models from SMART (23) and Pfam (which contains an overlapped TM2) databases.

Phylogenetic Tree Construction. All bacterial species for which at least one complete genome sequence was publicly available at the time the analysis began were included (≈500). Before building the phylogenetic tree, redundant species whose rRNA and ribosomal protein sequences were nearly identical [e.g., *B. pertussis* and *B. parapertussis* (24)], but which have different species names for phenotypic reasons, were identified and excluded. To do this, a multiple sequence alignment of the L5 ribosomal sequences was built and then used to construct a distance matrix using the JTT amino acid substitution matrix and

default parameters of Protdist in the Phylip package (25) (alignment method described later). Groups of sequences for which pair-wise distances among all members were less than 0.023 were identified. These groups corresponded to clades within a neighbor-joining tree built from the distance matrix using Neighbor from the Phylip package (25) with default parameters. Of these clades, only one member was chosen to represent the organism group in the final sequence set (Fig. S2). The 0.023 cutoff was chosen to ensure the exclusion of only highly related organisms because higher cutoffs resulted in sequence groups that partially span multiple clades. The final set of bacterial species included in the tree comprised 403 completely sequenced genomes plus the draft genomes of organisms studied in this work (*A. longum*, *T. primitia*, and *H. neapolitanus*). For the draft genomes, the contigs were subjected to the GeneMark gene finding program (26) to obtain the translated sequences.

Ribosomal protein sequences L3, L5, L11, L13, L14, S3, S5, S7, S8, S9, S11, and S17 were retrieved from the set of 406 genomes by using the Pfam domain models Ribosomal_L3, Ribosomal_L5_C, Ribosomal_L11, Ribosomal_L13, Ribosomal_L14, Ribosomal_S3_C, Ribosomal_S5_C, Ribosomal_S7, Ribosomal_S8, Ribosomal_S9, Ribosomal_S11, and Ribosomal_S17, respectively, and the HMMER software package (27). In rare instances in which multiple copies of a protein were encoded in the same genome, the copies were found to be identical or virtually identical and only one was collected for further analysis.

The 12 ribosomal sequence sets were individually aligned using the l-ins-i algorithm of MAFFT (28). The alignments were concatenated, and poorly conserved positions in the alignment were eliminated using Gblocks (29) with a maximum of 8 contiguous non-conserved positions, a 10-length minimum for a block, allowed positions with a gap for less than 50% of the sequences, and a 50% minimum number of sequences for a conserved or flanking position. The resulting alignment was used to build a LG+Γ_4+F maximum likelihood tree in PhyML 3.0 (30) with a sub-tree pruning and re-grafting topology search. For the genomes not analyzed by cryo-tomography, MCP sequences were identified using queries for the Pfam MCPsignal domain in the MIST database (31).

1. Ely B (1991) Genetics of *Caulobacter crescentus*. *Methods Enzymol* 204:372–384.
2. Eisenbach M, Adler J (1981) Bacterial cell envelopes with functional flagella. *J Biol Chem* 256:8807–8814.
3. Gluch MF, Typke D, Baumeister W (1995) Motility and thermotactic responses of *Thermotoga maritima*. *J Bacteriol* 177:5473–5479.
4. Komeili A, Li Z, Newman DK, Jensen GJ (2006) Magnetosomes are cell membrane invaginations organized by the actin-like protein MamK. *Science* 311:242–245.
5. Barbour AG (1984) Isolation and cultivation of Lyme disease spirochetes. *Yale J Biol Med* 57:521–525.
6. Angelakopoulos A, et al. (2002) Safety and shedding of an attenuated strain of Listeria monocytogenes with a deletion of actA/plcB in adult volunteers: a dose escalation study of oral inoculation. *Infect Immun* 70:3592–3601.
7. Leadbetter JR, Schmidt TM, Graber JR, Breznak JA (1999) Acetogenesis from H2 plus CO2 by spirochetes from termite guts. *Science* 283:686–689.
8. Cannon GC, Shively JM (1983) Characterization of homogenous preparation of carboxysomes from *Thiobacillus neapolitanus*. *Arch Microbiol* 134:52–59.
9. Vishniac W, Santer M (1957) The thiobacilli. *Bacteriol Rev* 21:195–213.
10. Dubochet J, et al. (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* 21:129–228.
11. Tivol W, Briegel A, Jensen GJ (2008) An improved cryogen for plunge freezing. *Microsc Micoanal* 14:375–379.
12. Iancu CV, et al. (2007) Electron cryotomography sample preparation using the Vitrobot. *Nat Protoc* 1:2813–2819.
13. Iancu C, et al. (2005) A 'flip-flop' rotation stage for routine dual-axis electron cryotomography. *J Struct Biol* 151:288–297.
14. Zheng QS, Braunfeld MB, Sedat JW, Agard DA (2004) An improved strategy for automated electron microscopy. *J Struct Biol* 147:91–101.
15. Suloway C, et al. (2009) Fully automated, sequential tilt-series acquisition with Leginon. *J Struct Biol* 167:11–8.
16. Mastronarde DA (1997) Dual-axis tomography: an approach with alignment methods that preserve resolution. *J Struct Biol* 120:343–352.
17. Hegerl R (1996) The EM program package: a platform for image processing in biological electron microscopy. *J Struct Biol* 116:30–34.
18. Wuichet K, Alexander RP, Zhulin IB (2007) Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. *Methods Enzymol* 422:1–31.
19. Alexander RP, Zhulin IB (2007) Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proc Natl Acad Sci USA* 104:2885–2890.
20. Bendtsen JD, Nielsen H, von Heinje G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
21. Cserzo M, Eisenhaber F, Eisenhaber B, Simon I (2004) TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics* 20:136–137.
22. Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res* 35:W429–W432.
23. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* 37:D229–D232.
24. Muller M, Hildebrandt A (1993) Nucleotide sequences of the 23s rRNA genes from *Bordetella pertussis*, *B parapertussis*, *B bronchiseptica* and *B avium*, and their implications for phylogenetic analysis. *Nucleic Acids Res* 21:3320.
25. Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* 5:164–166.
26. Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 33:W451–W454.
27. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
28. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511–518.
29. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577.
30. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
31. Ulrich LE, Zhulin IB (2007) MiST: a microbial signal transduction database. *Nucleic Acids Res* 35:D386–D390.
32. Ciccarelli FD, et al. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287.
33. Gupta RS (2000) The natural evolutionary relationships among prokaryotes. *Crit Rev Microbiol* 26:111–131.
34. Briegel A, et al. (2008) Location and architecture of the *Caulobacter crescentus* chemoreceptor array. *Mol Microbiol* 69:30–41.
35. Zhang P, et al. (2004) Direct visualization of receptor arrays in frozen-hydrated sections and plunge-frozen specimens of E. coli engineered to overproduce the chemotaxis receptor Tsr. *J Microsc* 216:76–83.

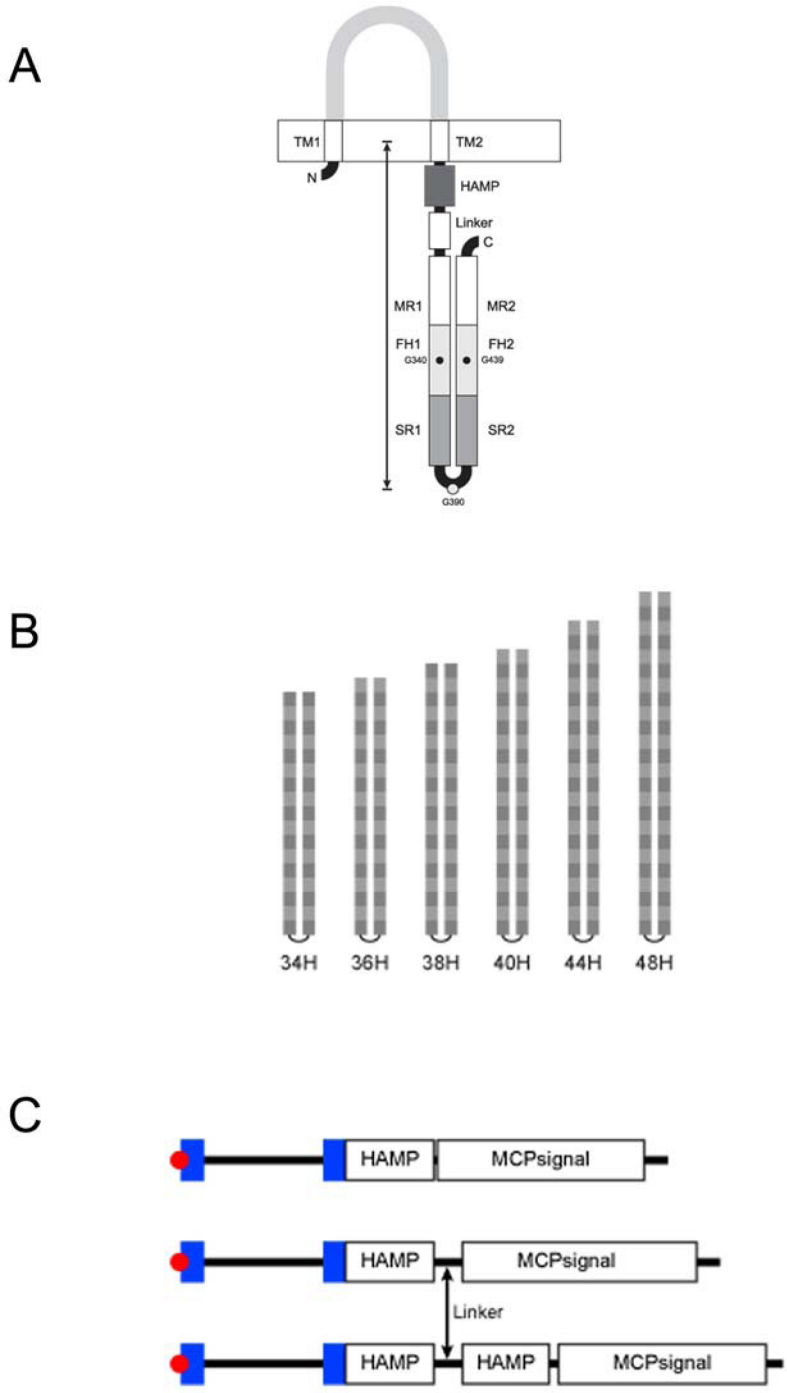

Fig. S1. MCP topology and architectural types. (*A*) Schematic representation of MCP topology type I. Two transmembrane regions (TM1 and TM2) anchor the receptor in the membrane. The extracellular (periplasmic) ligand-binding domain is shown in light gray. The cytoplasmic portion of the receptor consists of the HAMP domain, linker, and the signaling domain, which, in turn, is comprised of 3 sub-domains: methylation regions (MR1 and MR2, white), the flexible bundle sub-domain (FH1 and FH2, light gray), and the signaling sub-domain (SR1 and SR2, dark gray). In the *E. coli* Tsr receptor, Gly-340 and Gly-439 comprise the glycine ''hinge'' in the flexible bundle sub-domain, and the receptor hairpin is Gly-390. The ''cytoplasmic'' sequence length plotted on Fig. 3 was measured from the middle of TM2 to Gly-390 in the hairpin. (*B*) MCP signaling domain length classes (19) observed in the study. Light and dark gray blocks correspond to heptads (7 aa) for each class (i.e., there is a total of 34 heptads in the 34H MCP signaling domain). (*C*) The three different domain architectures of MCP topology type I receptors appearing in Fig. 3*B*. Red circles, predicted signal peptides; blue rectangles, predicted transmembrane regions; HAMP, predicted HAMP domains; linker, predicted linker region; MCPsignal, predicted signaling domains. The domain architecture at the top is most common. Receptor classes with the top, middle, and bottom domain architectures are identified by black, blue, and red labels in Fig. 3*B*, respectively.

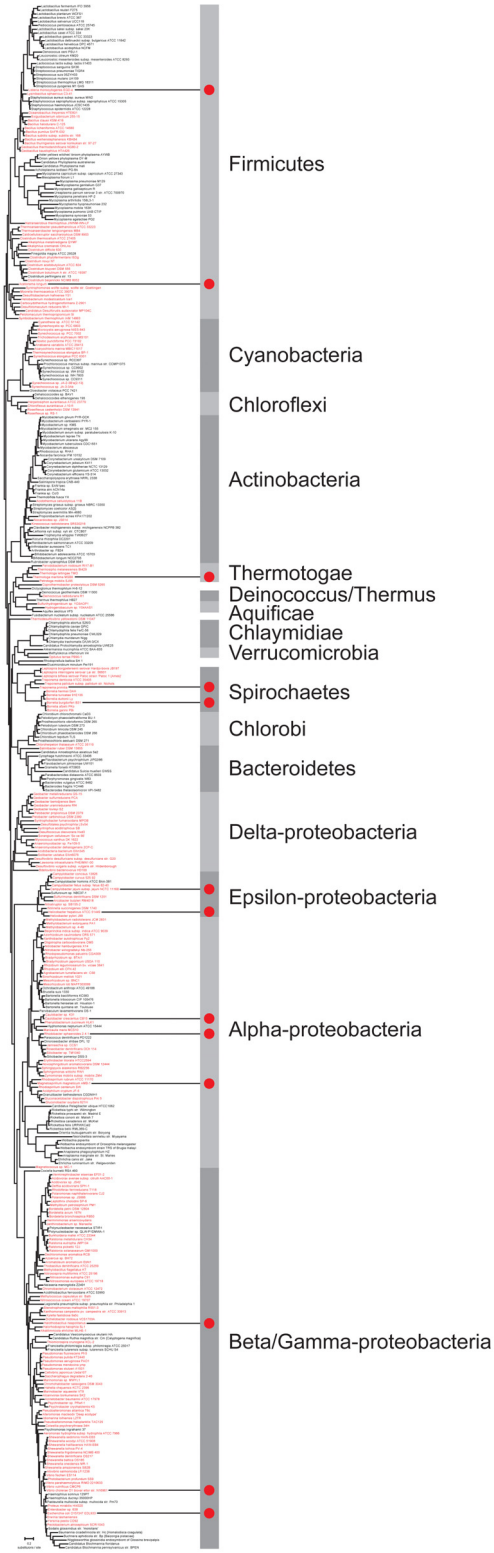

Fig. S2. Maximum likelihood phylogenetic tree of 406 representative bacteria. Organisms that encode one or more MCPs within their genomes are shown in red. Phyla are shown at the right (proteobacteria are further subdivided into classes) with the exception of unique organisms that are sole representatives of their phyla. Red circles and their associated lines show the placement of organisms examined in this study. The Firmicutes clade was used to root the tree, because it is currently considered to be the oldest phylum (32, 33). Note that the electronic (PDF) version of this figure located in *SI Appendix* can be enlarged sufficiently so that the individual species' names become readable.

Other Supporting Information Files

SI Appendix