Message from a human gut symbiont: sensitivity is a prerequisite for sharing

Jian Xu, Herbert C. Chiang, Magnus K. Bjursell and Jeffrey I. Gordon

Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, MO 63110, USA

Microbial genome sequencing projects are beginning to provide insights about the molecular foundations of human–bacterial symbioses. The intestine contains our largest collection of symbionts, where members of *Bacteroides* comprise ~25% of the microbiota in adults. The recently defined proteome of a prominent human intestinal symbiont, *Bacteroides thetaiotaomicron*, contains an elaborate environmental-sensing apparatus. This apparatus includes an unprecedented number of extracytoplasmic function (ECF) σ-factors, and a large collection of novel hybrid two-component systems composed of membrane-spanning periplasmic proteins with histidine kinase, phosphoacceptor, response regulator receiver and DNA-binding domains. These sensors are linked to the organism's large repertoire of genes involved in acquiring and processing dietary polysaccharides ('the glycobiome'). This arrangement illustrates how a successful symbiont has evolved strategies for detecting and responding to conditions in its niche so that it can sustain beneficial relationships with its microbial and human partners.

Symbiotic relationships shape all levels of biological organization, from single and multicellular life forms to complex communities and whole ecosystems in both oceanic and terrestrial realms [1–3]. These beneficial alliances are classified as: exosymbiotic when the partners remain external to one another; endosymbiotic when one partner resides inside the larger one but remains extracellular (e.g. the symbiosis between termites and their cellulose-metabolizing hindgut bacterial communities); or endocytobiotic (intracellular, as in the case of sponges belonging to the genus *Plakina* and cyanobacteria) [4]. Symbioses are typically predicated on syntropic interactions that allow utilization of a nutrient resource that neither partner can process alone. Microbial symbionts can function as 'keystone' species by producing effects on their ecosystems that are disproportionate to their abundance or biomass, and by greatly affecting biodiversity [5].

Despite the importance of animal–microbial symbioses, there have been only a few instances where the molecular mechanisms that forge and sustain these alliances have been deciphered experimentally [6,7]. One key question is how symbionts sense variations in their environments so that they can co-exist, co-adapt and co-evolve with their partners. The recently deciphered 6.26 Mb genome sequence of *Bacteroides thetaiotaomicron*, a Gram-negative obligate anaerobe, has provided insights about how this issue has been addressed by a very successful extracellular symbiont living in the densely populated distal intestinal ecosystem of humans [8].

The human gut microbiota

We acquire our consortia of microbes from the environment, beginning at the time of birth [9,10]. By adulthood, the number of bacteria that colonize our epidermal and mucosal surfaces is thought to exceed our population of human cells [11]. The largest community resides in our gut where 500–1000 species are assembled at densities estimated to reach 10^{11} per gram of luminal contents in the proximal colon, creating a bacterial 'nation' of 10–100 trillion citizens [12]. In this sense, we should not view ourselves as individuals, but rather as a highly diversified co-evolving collection of members of Eukarya, Bacteria and Archaea.

The phenotypic features of the majority of our bacterial partners are unknown because they cannot be cultivated *ex vivo*. However, new methods hold promise for overcoming this vexing problem [13], and for obtaining community-wide views of the metabolic processes that they support [14,15].

The collective size of the genomes of the 500–1000 species believed to reside in our intestine might be equivalent to the size of our own genome, and the number of bacterial genes in this 'microbiome' [16,17] could exceed the number of genes embedded in our human genome by a factor of 100 [18]. Because members of the microbiota provide metabolic traits that we ourselves have not fully evolved, including the ability to process nutrients that would be otherwise inaccessible [19], sequencing components of the microbiome can be viewed as a logical, albeit ambitious extension of the human genome project. This would offer an opportunity to obtain a more comprehensive view of ourselves as a life form, and at the same time provide insights into how symbiotic relationships are forged and how the environment affects the structure, evolution and function of microbial (and our own) genomes [20].

The Bacteroidetes phylum is one of the major lineages of Bacteria and arose very early during the course of evolution of this superkingdom [21]. Members of the *Bacteroides* genus account for ~25% of bacterial cells in the human intestine [11]. *B. thetaiotaomicron* is the first

Corresponding author: Jeffrey I. Gordon (jgordon@molecool.wustl.edu).

 0966-842X/$ - see front matter doi:10.1016/j.tim.2003.11.007

member of Bacteroidetes to be sequenced and the first predominant human symbiont for which the proteome has been deciphered [8]. *B. thetaiotaomicron* has numerous genes that allow it to acquire and process a broad range of dietary polysaccharides that we are not equipped to process on our own. For example, its 4779-member proteome contains 172 glycosylhydrolases that are predicted to be capable of degrading a broad range of natural glycosidic linkages, and 163 homologs of two outer membrane polysaccharide (starch)-binding proteins (SusC and SusD). Its highly evolved 'glycobiome' [8,22] is physically linked to an environmental-sensing apparatus that includes 50 extracytoplasmic function (ECF)-type σ^{70} factors and 33 hybrid two-component systems with a novel composite domain structure. Together, these elements provide an intriguing view of a sensory apparatus developed by a pre-eminent human symbiont.

Extracytoplasmic function (ECF)-type σ factors

The σ subunits of RNA polymerase (RNAP) complexes can be grouped into two structurally distinct families: σ^{70} (named after the 70 kDa primary σ-factor of *Escherichia coli*) and σ^{54} (named for the 54 kDa σ-factor involved in nitrogen regulation in *E. coli*). The majority of bacterial σ-factors belong to the σ^{70} family, and all characterized bacteria appear to contain at least one family member [23].

ECF σ-factors are the largest and most divergent group in the σ^{70} family [24,25]. They can be distinguished from other family members by their small size (typically 20–30 kDa), unique domain composition, the promoter sequences they recognize, and their frequent regulation by specific anti σ-factors [25]. Previously characterized ECF σ-factors control a variety of functions, including iron-uptake and heat shock responses in *E. coli* [26,27], alterations in outer membrane protein (OMP) synthesis in *Photobacterium* spp. that coincide with changes in osmolarity, barometric pressure or temperature [28], and the cellular adjustments that are made when disulphide stress is experienced by the ubiquitous soil bacterium *Streptomyces coelicolor* [29].

A prototypic pathway of gene regulation by ECF σ-factors involves receipt of an environmental stimulus followed by release of the protein from a membrane-tethered cognate anti σ-factor so that the ECF σ can interact with RNAP [24]. A well-studied example is *E. coli* σ^{E} and its anti σ-factor, known as RseA. A crystal structure of a complex of these two proteins has been solved to a resolution of 2 Å [30] (Figure 1). The family members of σ^{70} have four conserved regions that are numbered 1 to 4 [31]. Regions 2 and 4 are well conserved in all σ^{70} family members, whereas regions 1 and 3 are absent in ECF σ-factors. The factor σ^{E} consists of an N-terminal domain that corresponds closely to conserved region 2, followed by a 26-residue flexible linker and a C-terminal domain that represents conserved region 4. RseA is an inner membrane protein. The N-terminal cytoplasmic domain interacts with σ^{E} where it is sandwiched between regions 2 and 4, whereas the C-terminal periplasmic domain interacts with other regulators (RseB and RseC) [32,33] (Figure 1). RseA prevents σ^{E} from binding to the core RNAP [30]. This inhibition is relieved through proteolytic degradation of RseA [34–36]. The proteolytic cascade is initiated by environmental factors that result in misfolding of proteins in the periplasm or outer membrane (e.g. heat shock) and involves DegS-mediated removal of the C-terminal periplasmic domain of RseA [34,35], YaeL-directed cleavage

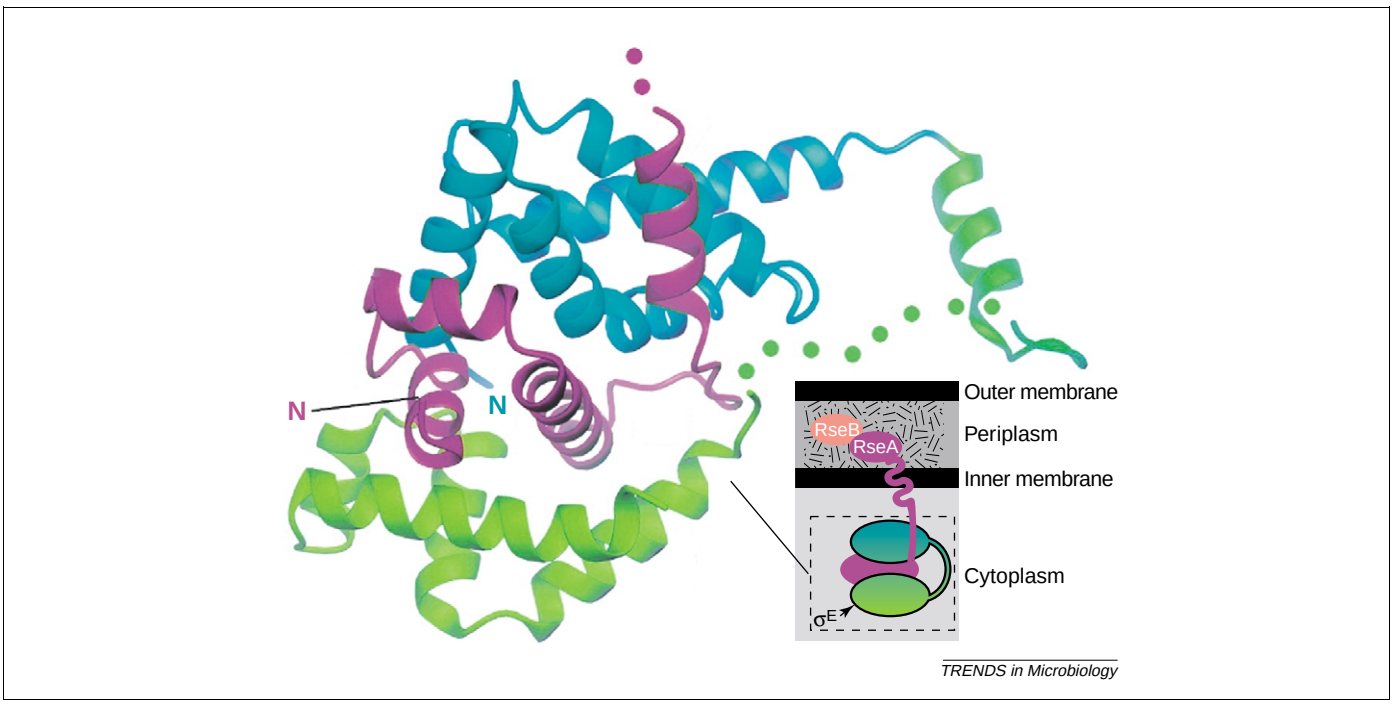

Figure 1. Complex formed between the *Escherichia coli* extracytoplasmic function (ECF) σ-factor σ^{E} and its RseA anti σ-factor. Ribbon diagram of the 2.0 Å crystal structure of σ^{E} complexed with the N-terminal cytoplasmic domain of RseA (PDB accession number **1OR7**). σ^{E} contains two of the four domains represented in σ^{70} family members: an N-terminal region (aquamarine, conserved region 2) separated by a flexible linker from a C-terminal region (green, conserved region 4). Green dots denote a portion of the C-terminal region of σ^{E} that has been omitted from the ribbon diagram. RseA contains an N-terminal cytoplasmic domain (purple), an inner membrane-spanning domain, and a C-terminal periplasmic region that interacts with other regulators, including RseB (see inset). Adapted from [30].

Table 1. Top bacterial genomes in terms of their ECF σ-factor content[a]

Organism	Genome size (Mb)	Total number of σ-factors	Number of ECF σ-factors	Refs
Bacteroides thetaiotaomicron	6.26	54	50	[8]
Streptomyces avermitilis	9.03	62	47	[66]
Streptomyces coelicolor	8.67	68	45	[67]
Nitrosomonas europaea	2.81	29	22	[68]
Pseudomonas aeruginosa	6.26	25	19	[69]
Bacillus anthracis	5.23	26	15	[70]

[a]Data on the number of σ-factors are based on annotations associated with each genome.

within or near the transmembrane domain of RseA [37–39], and release of the σ^E 'RseA–cyto' complex into the cytosol (RseA–cyto, the N-terminal region of RseA, is sufficient for anti-σ activity [32,33]). Additional degradation then occurs, possibly by ClpXP [40], so that σ^E can be incorporated into the RNAP holoenzyme [25].

B. thetaiotaomicron contains the largest proportion (and absolute number) of ECF σ-factors among the 125 species of Bacteria and Archaea for which complete genome sequences have been deposited in public databases as of July 12, 2003 (Table 1). Of the 54 σ-factors identified in *B. thetaiotaomicron*, BLAST searches indicate that 48 have E values $<10^{-6}$ to known ECF σ-factors in SwissProt, whereas two others (BT2169 and BT2562) have homology to a putative *Bacteroides fragilis* ECF σ-factor known as RpoE (GenBank accession number **18033128**) (22% and 28% global amino acid sequence similarity, respectively). *B. thetaiotaomicron* ECF σ-factors contain 147–202 amino acids. The one exception, BT2778, contains 549. Multiple sequence alignments [41] indicate that all have conserved domains at their N- and C-termini, corresponding to regions 2 and 4 (Figure 2). These conserved domains include 'sub-regions' of σ^{70} family members involved in binding to the core RNAP, recognition of -10 and -35 promoter elememts, as well as melting of the promoter [25].

Twenty-five of the 50 genes that encode these proteins are located next to open reading frames (ORFs) specifying putative anti σ-factors [8]. These anti σ-factors form a novel paralogous group, all with best hits to *B. fragilis* FrrF (GenBank accession number **18033129**). FrrF, annotated as a 'putative transmembrane sensor protein', is located in a portion of the *B. fragilis* genome (http://www.sanger.ac.uk/Projects/B_fragilis/) that contains a cluster of regulatory genes (Frr, 'fragilis regulatory region'; Coyne and Comstock, unpublished). The 25 candidate *B. thetaiotaomicron* anti σ-factors, as well as FrrF, have weak sequence similarity to *E. coli* FecR (GenBank accession number **1790745**; E values range from 10^{-3} to 10^{-8}; whole length sequence similarity <20%). *E. coli* FecR, together with FecA (OMP) and FecI (ECF σ), regulates expression of genes involved in ferric citrate transport (*fecABCDE*; [42]).

Several additional findings support the notion that these 25 *B. thetaiotaomicron* proteins are anti σ-factors. Known anti σ-factors are typically membrane-associated [43]. According to PSORT (http://psort.nibb.ac.jp/) [44], 24 of the 25 proteins are targeted to the inner membrane;

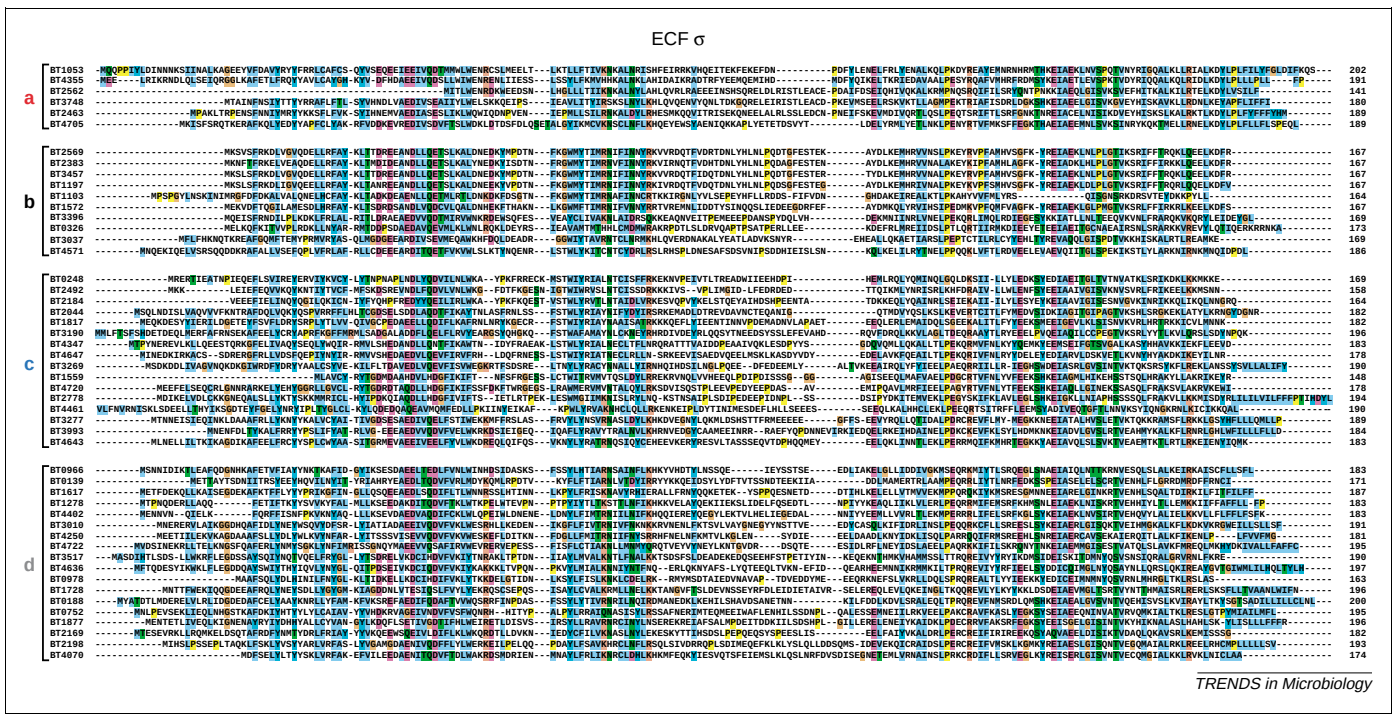

Figure 2. *Bacteroides thetaiotaomicron* extracytoplasmic function (ECF) σ-factors. This multiple sequence alignment was generated using Clustal W. BT numbers refer to gene designations in the *B. thetaiotaomicron* genome (GenBank accession number **AE015928**). Four subgroups were defined by phylogenetic analysis (labeled a–d). Conserved residues are highlighted.

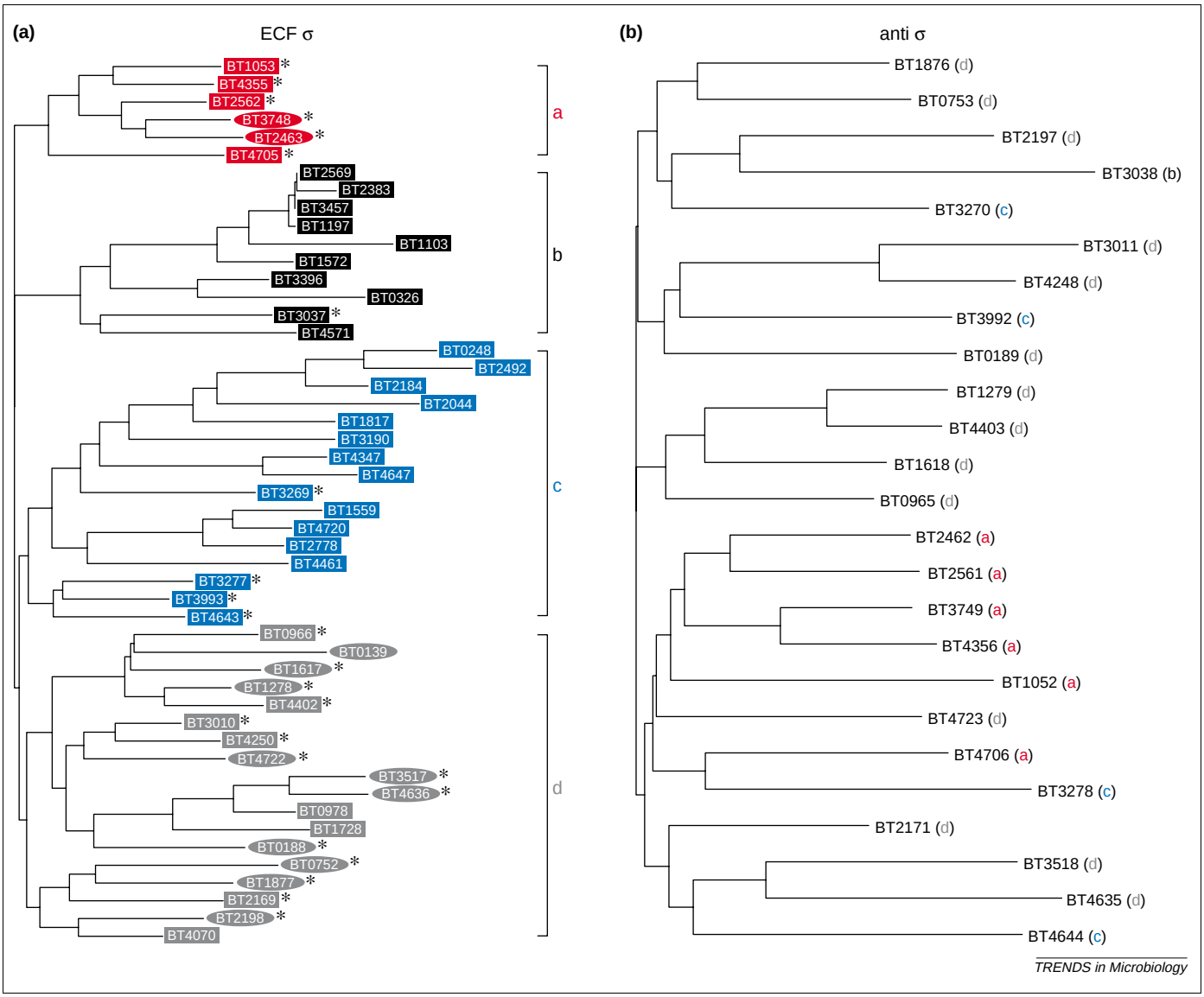

Figure 3. *Bacteroides thetaiotaomicron* extracytoplasmic function (ECF) σ-factors. **(a)** Phylogenetic tree of the 50 ECF σ-factors showing four identified subgroups. Genes with an asterisk are paired (physically linked) with a gene encoding a putative anti σ-factor. Those highlighted with an oval-shaped color are associated with 12 gene clusters that contain downstream SusC–D homologs and glycosyltransferases. **(b)** Phylogenetic tree of the 25 putative anti σ-factors present in the proteome. The letters in parenthesis indicate the subgroup identity of the physically linked ECF σ-factor gene.

BT4248 is the only exception and is predicted to be cytoplasmic. ECF σ-factors are typically located upstream of a co-transcribed anti σ [43]. Twenty-three of the 25 putative anti σ-factor ORFs are located immediately downstream of an ECF σ, and 18 of these 23 are positioned on the same DNA strand. The other two anti σ-factors (BT2171 and BT4248) are separated from an ECF σ by a single ORF encoding a hypothetical protein.

Multiple sequence alignments using Clustal W [41] and Phylip [45] indicate that the 50 *B. thetaiotaomicron* ECF σ-factors do not form a single distinct group when compared with other ECF σ-factors, and have diverged from one another to form four subgroups (designated a–d in Figure 3a). All members of subgroup a (6 out of a total of 6) and the majority of subgroup d members (14 out of 18) are paired with anti σ-factors in the bacterial chromosome, in contrast to members of subgroup b and c (1 of 10 and 4 of 16, respectively). Multiple sequence alignments also disclose that the 6 anti σ-factors paired with subgroup a ECF σ-factors form a distinct cluster in the anti σ-factor phylogenetic tree (Figure 3b). These observations provide clues about possible functional relationships that might have evolved between subsets of ECF σ and anti σ-factors.

ECF σ and anti σ-factors present in the *B. thetaiotaomicron* genome appear to play roles in regulating components of its glycobiome, perhaps imparting the ability to opportunistically retrieve and metabolize different types of glycans depending upon their availability in the host niche. All but one of the ECF and anti σ-factor gene pairs is positioned immediately upstream of ORFs that encode homologs of the polysaccharide-binding SusC OMP. *B. thetaiotaomicron* contains 12 gene clusters with a conserved modular structure consisting of an ECF σ, followed by an anti σ (11 of 12 cases), SusC and SusD homologs, one or more glycosylhydrolases, plus other enzymes involved in carbohydrate metabolism [8]. None of these clusters is associated with an upstream ORF specifying other classes of transcriptional regulators.

The remaining 25 ECF σ-factor genes that are not physically linked to a putative anti σ-factor are positioned immediately upstream of genes with a variety of functions: SusC homologs in three cases; metabolic enzymes in 14 cases (phosphoesterase, oxidoreductase, helicase, methyltransferase and β-glucosidase among others); a putative transport-related protein in one case; a putative transmembrane sensor protein in another case; and hypothetical proteins in six other cases [8]. These features are not apparent in the genomes of the few sequenced members of other genera represented in the distal adult human intestinal microbiota [22].

Adaptive foraging behavior stabilizes food webs [46]. Therefore, a highly evolved capacity to accurately sense ('taste') the intestine's glycan landscape and mobilize a response that allows these nutrients to be harvested should benefit (i) *B. thetaiotaomicron* itself, (ii) other members of the gut microbial community able to scavenge oligosaccharides generated by the 'digestive system' of *B. thetaiotaomicron* (for example, *Bifidobacterium longum* [22,47]), and (iii) the human consumer of otherwise indigestible polysaccharides who, in this syntropic relationship, obtains 10–15% of his or her daily calories from short chain fatty acids (acetate, butyrate and propionate) generated through microbial fermentation of oligosaccharides [48].

The σ-factors and their anti σ-factors have evolved to recognize a wide variety of signals, and might operate through several mechanisms [43]. Known anti σ-factors can display remarkable specificity; for example, they are able to select a single σ-factor from a cellular milieu that can be simultaneously populated by many σ-factors that have significant sequence similarity [43]. *B. thetaiotaomicron* provides an attractive model system for dissecting how environmental cues, for example, the presence or absence of specific types of carbohydrates, are sensed and transduced via ECF-type σ and anti σ-factors to the genome to modulate bacterial physiology. The diversity of signal-sensing and transduction by ECF σ-factors might be largely achieved through operation of their structurally diverse cognate anti σ-factors [30]. Therefore, the selectivity of each of the 50 ECF σ-factors in *B. thetaiotaomicron* for the organism's 25 candidate anti σ-factors, and the degree to which such selectivity is legislated by their availability versus their specificity for one another needs to be determined. The issue of availability can be addressed in part by DNA microarray-based profiling of the expression of ECF σ-factor genes and their paired (and unpaired) anti σ-factors during growth *ex vivo* under defined environmental conditions. The issue of selectivity represents an intriguing and challenging problem for *in silico* and experimental analyses.

Hybrid two-component systems

Like ECF σ and anti σ-factors, two-component systems function in prokaryotes as regulators of various environmental signaling transduction pathways [49]. The prototypic two-component system consists of a membrane-localized histidine kinase sensor containing histidine kinase and phosphoacceptor domains (HATPase_c and HisKA in the Pfam database; http://pfam.wustl.edu/), and a response regulator containing a response regulator receiver domain (response_reg) and a signal output domain [50–57] (Figure 4).

The majority of known response regulators are transcription factors with output domains that can be divided into three major sub-families [58] based on their DNA-binding domains: OmpR–PhoB winged-helix domains (trans_reg_c in Pfam), NarL–FixJ four-helix domains (GerE in Pfam) and NtrC ATPase-coupled transcription factors (HTH_8 in Pfam). The trans_reg_c family appears to be the largest among the 29 response regulators of *B. thetaiotaomicron*. This organism also contains 54 genes encoding traditional sensor kinases, 20 of which are directly linked to ORFs encoding response regulators.

A 'traditional' hybrid histidine kinase consists of a single polypeptide with a histidine kinase domain, a phosphoacceptor domain and a response regulator receiver domain [50] (Figure 4). None of the previously annotated hybrid histidine kinases in either prokaryotes or eukaryotes contains a DNA-binding domain; their interactions with the genome are achieved through separate response regulators in a multi-component phospho-relay system (Figure 4). The 125 publicly available sequenced prokaryotic genomes encode 272 proteins with HATPase_c, HisKA and response_reg domains (E value $<10^{-3}$). Only a small number of additional Pfam domains are evident in these hybrid histidine kinases; those occurring more than once are PAC, PAS, HAMP, GAF, Hpt, SBP_bac_3, CBS, SPNTR and Cache, in decreasing order of frequency (domain definitions can be found at Pfam text search http://pfam.wustl.edu/textsearch.shtml). None of these domains are known to interact with DNA. By contrast, *B. thetaiotaomicron* contains 32 novel hybrid histidine kinases with a DNA-binding domain. These large inner membrane proteins are composed of an N-terminus predicted to be periplasmic (according to PSORT), followed by one to five transmembrane segments, which can be predicted using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html); HisKA, HATPase_c and response_reg domains; and one or two AraC-type helix-turn-helix domains (HTH_AraC) (Figure 4). It is the presence of this HTH_AraC DNA-binding domain (http://pfam.wustl.edu/cgi-bin/getdesc?name = HTH_AraC) [59] that distinguishes the hybrid histidine kinases of *B. thetaiotaomicron* from previously described models of two-component systems (Figure 4).

These newly described hybrid histidine kinases constitute a major component of the network of 'two-component' regulators in *B. thetaiotaomicron* (Table 2), and can be grouped into several subsets on the basis of multiple sequence alignments (Figure 5). As with ECF σ-factors, many of these genes are positioned next to components of the glycobiome (Figure 5). Nineteen are adjacent to genes encoding oligo or polysaccharide hydrolases, three are adjacent to sulfatases, and one is deposited next to a heparitin sulfate lyase. These sulfatases and the lysase are important for degradation of mucopolysaccharides present in the mucus layer that overlies the intestinal epithelium, and provides a nutrient-rich pasture for *B. thetaiotaomicron* grazing. Of course, adjacency does not establish that these genes are under the control of

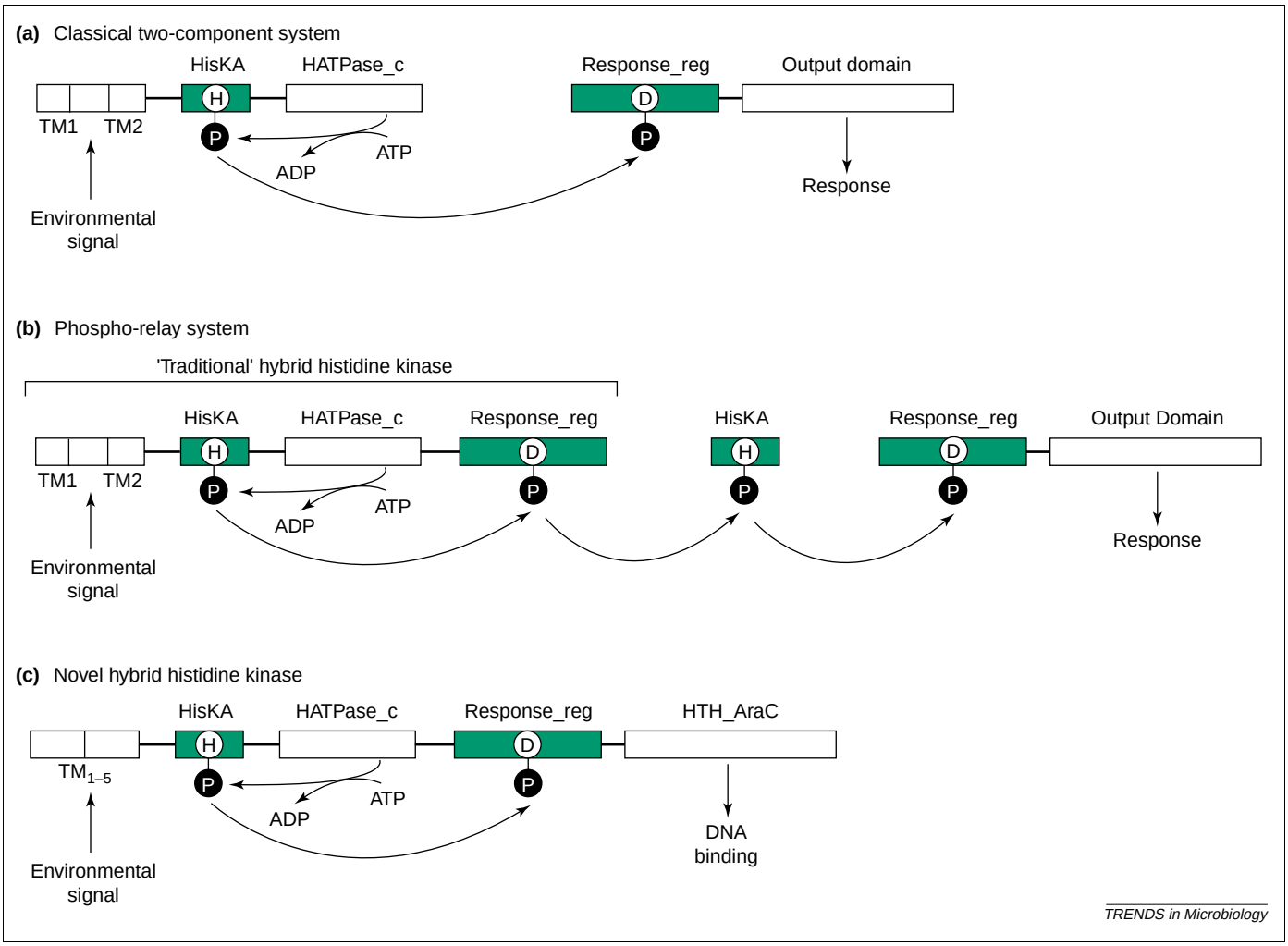

Figure 4. Characterization of novel hybrid two-component systems in *Bacteroides thetaiotaomicron*. Schematic overview of two-component systems showing Pfam domains (HATPase_c, histidine kinase domain; HisKA, phosphoacceptor domain; response_reg, response regulator receiver domain; HTH_AraC, AraC-type DNA binding domain; TM, transmembrane domain). **(a)** Classical two-component system [56]. **(b)** Phospho-relay system consisting of a 'traditional' hybrid histidine kinase, one or more phosphotransfer proteins, and a response regulator [56]. **(c)** Novel hybrid histidine kinases, discovered in *B. thetaiotaomicron*, that incorporate all of the functional components of a two-component system into a single polypeptide [22]. The number of TM domains varies from 1 to 5 among these novel hybrid histidine kinases. Adapted from [58].

these novel hybrid histidine kinases. This notion must be tested directly by gene disruption and whole genome transcriptional profiling of isogenic wild type and mutant strains.

Table 2. Grouping of all sensor histidine kinases present in the *Bacteroides thetaiotaomicron* proteome[a]

Pfam domain				Number of Proteins
HisKA	HATPase_c	Response_reg	HTH_AraC	
X	X	X	X	32
X	X	X		8
X	X		X	1
X	X			27
	X			7
X				2
no conserved domains				8

[a]Based on the representation of four Pfam core domains. Proteins without any conserved domains have significant global sequence similarity to known sensor kinases in two-component systems (E value $< 10^{-6}$ as defined by BLASTP).

The unique domain structure and abundant representation of hybrid histidine kinases in *B. thetaiotaomicron* raises several questions. What advantages do they provide in detecting and transducing environmental signals to the genome compared with classical two-component systems? The presence of HATPase_c, HisKA, response_reg and HTH_AraC domains allows these 32 proteins to incorporate all of the features that are used by a two-component system to directly couple receipt of an environmental signal to regulation of gene expression. In contrast to two separate proteins that operate through a simple, two-partner phospho-transfer scheme (common in bacteria), or a multiple-partner 'phosphor-relay' cascade (more prevalent in eukaryotes [58,60]), these novel hybrid histidine kinases might be designed to constrain signal amplification, modularity or cross-talk between two-component systems [61]. If a central challenge to a microbial symbiont living in a densely populated distal intestinal niche is to discriminate between diverse environmental chemical entities that are present at high concentrations, then evolving and expanding a unique family of signaling

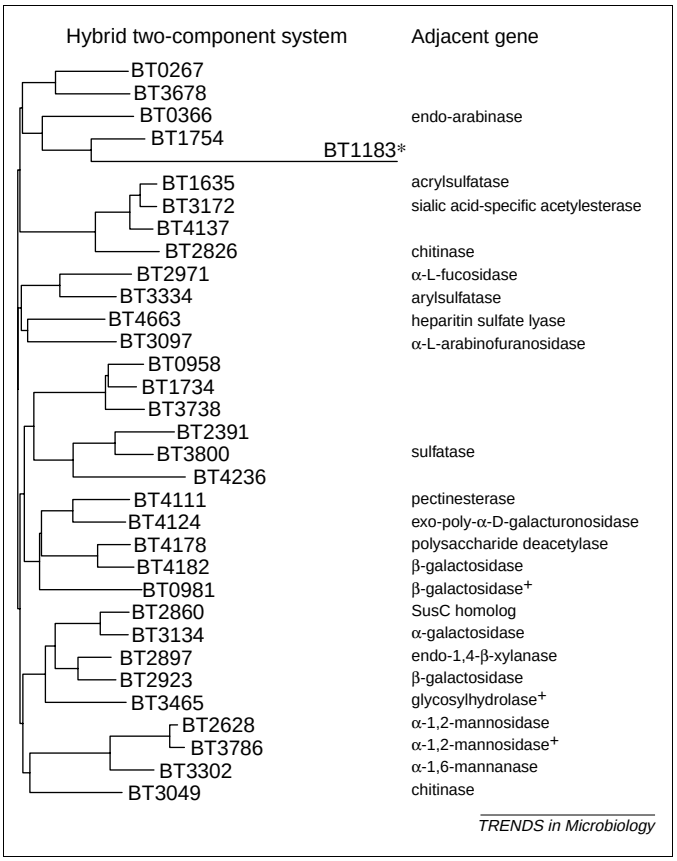

Figure 5. Phylogenetic analysis of the organism's 33 novel hybrid histidine kinases, based on Clustal W. The predicted functions of neighboring genes are shown. Abbreviations: *, contains a glycos_transf domain in place of a HTH_AraC domain; +, the gene encoding the hybrid histidine kinase is separated from the listed adjacent gene by an open reading frame (ORF) specifying a hypothetical protein.

molecules that sacrifice signal amplification for precision and specificity might impart a significant advantage. Intriguingly, *B. fragilis*, which is closely related to *B. thetaiotaomicron* but only a minor component of the normal human gut microbiota, contains disproportionately fewer hybrid histidine kinases in its proteome (a total of 8 in a genome that is only 17% smaller than *B. thetaiotaomicron*). These considerations raise the question as to what signals are recognized by these hybrid proteins, and whether they are directly related to a key metabolic mission of this symbiont: namely, to acquire and process glycans available in its ecosystem for its own use and for distribution to other 'colleagues' in the microbial community and to the host [22].

The mechanism by which this unique family of signaling molecules regulates gene transcription is also unclear. Is proteolytic processing required to release a portion of the protein that contains the DNA-binding domain so that it can interact with *cis*-acting regulatory elements in target genes? If so, what is the nature of the proteases and is their expression, trafficking and/or turnover regulated by environmental signals?

Finally, because cross-talk between ECF σ-factors and classical two-component systems has been documented in other organisms [62], what is the degree of interconnection among the ECF and anti σ-factor regulators, the classical two-component systems and the novel hybrid histidine kinases of *B. thetaiotaomicron*?

Prospects

The availability of genetic systems for manipulating *B. thetaiotaomicron* [63], the relative ease of culturing this aero-tolerant anaerobe, and a complete genome sequence [8] sets the stage for direct genetic tests of how components of its environmental-sensing apparatus influence its physiology and symbiosis with humans. Developing methods for rapid, broad-based phenotyping of panels of engineered isogenic mutants lacking each of the hybrid two-component systems, or each of the ECF σ-factors (or anti σ-factors), will be crucial for this effort. 'Phenotype microarrays' [64,65] provide one tool for interpreting (i) the results of genetic tests of the roles of regulatory network components, and (ii) datasets of whole genome transcriptional profiles obtained from bacterial mutants grown under defined environmental conditions.

As with marriages between humans, understanding how symbionts have evolved strategies for effective 'listening' (sensing) is a first step in understanding how these relationships are able to endure and provide mutual support and benefit in the face of inevitable changes in their home or work environments.

Acknowledgements

We thank Justin Sonnenburg for helpful comments concerning this manuscript. Work from the author's laboratory described in this report was supported in part by a grant from the National Institutes of Health (DK30292).

References

1 Boucher, D.H. (1985) *The Biology of Mutualism: Ecology and Evolution*, Oxford University Press
2 Ahmadjian, V. and Paracer, S. (1986) *Symbiosis: An Introduction to Biological Associations*, University Press of New England, London
3 Margulis, L. and Sagan, D. (2002) *Acquiring Genomes, a Theory of the Origins of Species*, Basic Books, NY
4 Nardon, P. and Charles, H. (2002) Morphological aspects of symbiosis. In *Symbiosis: Mechanisms and Model Systems* (Seckbach, J., ed.), pp. 15–44, Kluwer Academic Publishers
5 Zook, D.P. (2002) Prioritizing symbiosis to sustain biodiversity: are symbionts keystone species? In *Symbiosis: Mechanisms and Model Systems* (Seckbach, J., ed.), pp. 5–12, Kluwer Academic Publishers
6 McFall-Ngai, M.J. (2002) Unseen forces: the influence of bacteria on animal development. *Dev. Biol.* 242, 1–14
7 Gualtieri, G. and Bisseling, T. (2000) The evolution of nodulation. *Plant Mol. Biol.* 42, 181–194
8 Xu, J. *et al.* (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299, 2074–2076
9 Mackie, R.I. *et al.* (1999) Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* 69, 1035S–1045S
10 Favier, C.F. *et al.* (2002) Molecular monitoring of succession of bacterial communities in human neonates. *Appl. Environ. Microbiol.* 68, 219–226
11 Savage, D.C. (1977) Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* 31, 107–133
12 Berg, R.D. (1996) The indigenous gastrointestinal microflora. *Trends Microbiol.* 4, 430–435
13 Rappe, M.S. *et al.* (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630–633
14 Beja, O. *et al.* (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* 415, 630–633
15 Orphan, V.J. *et al.* (2001) Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* 293, 484–487
16 Relman, D.A. and Falkow, S. (2001) The meaning and impact of the human genome sequence for microbiology. *Trends Microbiol.* 9, 206–208

17 Lederberg, J. (2000) Infectious history. *Science* 288, 287–293
18 Hooper, L.V. and Gordon, J.I. (2001) Commensal host-bacterial relationships in the gut. *Science* 292, 1115–1118
19 Hooper, L.V. *et al.* (2002) How host–microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* 22, 283–307
20 Stahl, D.A. and Tiedje, J. (2002) Microbial ecology and genomics: a crossroads of opportunity. *Critical Issues Colloquia, American Society of Microbiology*
21 Garrity, G. (2002) *Bergey's Manual of Systematic Bacteriology* (Vol. 1), Springer Verlag, p. 58
22 Xu, J. and Gordon, J.I. (2003) Honor thy symbionts. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10452–10459
23 Wosten, M.M. (1998) Eubacterial sigma-factors. *FEMS Microbiol. Rev.* 22, 127–150
24 Helmann, J.D. (2002) The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* 46, 47–110
25 Paget, M.S. and Helmann, J.D. (2003) The sigma70 family of sigma factors. *Genome Biol.* 4, 203
26 Erickson, J.W. and Gross, C.A. (1989) Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev.* 3, 1462–1471
27 Enz, S. *et al.* (1995) Transcription of the region encoding the ferric dicitrate-transport system in *Escherichia coli*: similarity between promoters for *fecA* and for extracytoplasmic function sigma factors. *Gene* 163, 13–18
28 Chi, E. and Bartlett, D.H. (1995) An *rpoE*-like locus controls outer membrane protein synthesis and growth at cold temperatures and high pressures in the deep-sea bacterium *Photobacterium* sp. strain SS9. *Mol. Microbiol.* 17, 713–726
29 Paget, M.S. *et al.* (2001) Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol. Microbiol.* 42, 1007–1020
30 Campbell, E.A. *et al.* (2003) Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell* 11, 1067–1078
31 Lonetto, M. *et al.* (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *J. Bacteriol.* 174, 3843–3849
32 De Las Penas, A. *et al.* (1997) The sigmaE-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol. Microbiol.* 24, 373–385
33 Missiakas, D. *et al.* (1997) Modulation of the *Escherichia coli* sigmaE (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol. Microbiol.* 24, 355–371
34 Ades, S.E. *et al.* (1999) The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes Dev.* 13, 2449–2461
35 Alba, B.M. *et al.* (2001) degS (hhoB) is an essential *Escherichia coli* gene whose indispensable function is to provide sigma (E) activity. *Mol. Microbiol.* 40, 1323–1333
36 Ades, S.E. *et al.* (2003) Regulation of the alternative sigma factor sigma(E) during initiation, adaptation, and shutoff of the extracytoplasmic heat shock response in *Escherichia coli*. *J. Bacteriol.* 185, 2512–2519
37 Kanehara, K. *et al.* (2001) Characterization of the *yaeL* gene product and its S2P-protease motifs in *Escherichia coli*. *Gene* 281, 71–79
38 Alba, B.M. *et al.* (2002) DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response. *Genes Dev.* 16, 2156–2168
39 Kanehara, K. *et al.* (2002) YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. *Genes Dev.* 16, 2147–2155
40 Flynn, J.M. *et al.* (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* 11, 671–683
41 Thompson, J.D. *et al.* (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
42 Braun, V. *et al.* (2003) Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Curr. Opin. Microbiol.* 6, 173–180
43 Hughes, K.T. and Mathee, K. (1998) The anti-sigma factors. *Annu. Rev. Microbiol.* 52, 231–286
44 Nakai, K. and Kaneshisa, M. (1991) Expert system for predicting protein localization sites in Gram-negative bacteria. *Proteins* 11, 95–110
45 Felsenstein, J. (1989) PHYLIP – Phylogeny Inference Package (version 3.2). *Cladistics* 5, 164–166
46 Kondoh, M. (2003) Foraging adaptation and the relationship between food-web complexity and stability. *Science* 299, 1388–1391
47 Schell, M.A. *et al.* (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14422–14447
48 Bergman, E.N. (1990) Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70, 567–590
49 Stock, A.M. *et al.* (2000) Two-component signal transduction. *Annu. Rev. Biochem.* 69, 183–215
50 Wolanin, P.M. *et al.* (2002) Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol.* 3 REVIEWS3013 (http://genomebiology.com)
51 Mizuno, T. (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.* 4, 161–168
52 Rodrigue, A. *et al.* (2000) Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol.* 8, 498–504
53 Throup, J.P. *et al.* (2000) A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Mol. Microbiol.* 35, 566–576
54 Koretke, K.K. *et al.* (2000) Evolution of two-component signal transduction. *Mol. Biol. Evol.* 17, 1956–1970
55 Galperin, M.Y. *et al.* (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.* 203, 11–21
56 Mizuno, T. *et al.* (1996) Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803. *DNA Res.* 3, 407–414
57 Wang, L. *et al.* (2002) Genomic analysis of protein kinases, protein phosphatases and two-component regulatory systems of the cyanobacterium *Anabaena* sp. strain PCC 7120. *FEMS Microbiol. Lett.* 217, 155–165
58 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
59 Gallegos, M.T. *et al.* (1993) The XylS/AraC family of regulators. *Nucleic Acids Res.* 21, 807–810
60 Santos, J.L. and Shiozaki, K. (2001) Fungal histidine kinases. *Sci. STKE.* RE1 (http://stke.sciencemag.org)
61 Bijlsma, J.J.E. and Groisman, E.A. (2003) Making informed decisions: regulatory interactions between two-component systems. *Trends Microbiol.* 11, 359–366
62 Paget, M.S. *et al.* (1999) A putative two-component signal transduction system regulates sigmaE, a sigma factor required for normal cell wall integrity in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* 33, 97–107
63 Salyers, A.A. *et al.* (2000) Starting a new genetic system: lessons from bacteroides. *Methods* 20, 35–46
64 Bochner, B.R. (2003) New technologies to assess genotype-phenotype relationships. *Nat. Rev. Genet.* 4, 309–314
65 Zhou, L. *et al.* (2003) Phenotype microArray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *J. Bacteriol.* 185, 4956–4972
66 Ikeda, H. *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* 21, 526–531
67 Bentley, S.D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147
68 Chain, P. *et al.* (2003) Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J. Bacteriol.* 185, 2759–2773
69 Stover, C.K. *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406, 959–964
70 Read, T.D. *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423, 81–86