

Supporting Information

Singh *et al.* 10.1073/pnas.0709764105

SI Methods

Phenotype Data. Phenotypes for 209 bacterial species with fully sequenced genomes were collected from the literature, including Bergey's Manual (1), NCBI Microbial Genomes (www.ncbi.nlm.nih.gov/genomes/lproks.cgi), and TIGR Genome Properties (2). We gathered qualitative information on habitat, metabolism, electron source, electron donor, and scored whether the species had a number of phenotypes: motility, spore-formation, transformability for DNA; as well as whether it is anaerobic, an extremophile, or Gram-positive (Table S1).

Building Phylogenetic Profiles. To build phylogenetic profiles of stress responses, we used 61 genes in chemotaxis, 153 genes in spore-formation, and 62 genes in competence (Tables S2–S4). These genes were chosen after extensive literature search of each stress response, including separate experimental evidence (genetic, biochemical, or high-throughput expression data) for each gene in the model organisms *B. subtilis* for endospore formation and chemotaxis, and *B. subtilis* and *H. influenzae* for competence (33 and 29 genes, respectively). For comparison, an additional complete analysis was performed with *E. coli* chemotaxis genes as the starting set and the statistical results were confirmed to be qualitatively similar to those reported here (data not shown). DNA and amino acid sequences for all genes were retrieved from the MicrobesOnline website on October 8, 2006 and their orthologs were identified by a three-way bidirectional best hit algorithm as previously described (3), with the additional constraint that the sequence alignment coverage had to be at least 75% of the length of both genes. For phylogenetic profiles, a species was marked as having a gene if it had at least one ortholog, but possibly also multiple paralogs, of the gene. Ortholog sets for transcriptional regulators and histidine kinases, which are known to have highly conserved domains (4, 5), were manually curated to remove spurious hits by examining the phylogenetic tree for each ortholog set. Nucleotide and amino acid alignments were performed using Muscle (6) with $\text{max_iters} = 3$ and $\text{diags} = 1$ (paralogs were discarded before the alignment step). Phylogenetic trees were built from aligned amino acid sequences using PhyML (7) with default optimization parameters and 100 bootstrap replicates, and visualized on the iTOL tree viewer (<http://itol.embl.de>).

Criteria for Module Classification. To decide the optimal number of clusters for gene dendrograms, we calculated the mean silhouettes (8) for cuts along each gene tree producing two to 10 clusters, and chose the partition with maximal mean silhouette over all clusters. The same number of genes randomly chosen (1,000 iterations) from either the *B. subtilis* genome (for sporulation and chemotaxis) or a proportional mixture of the *B. subtilis* and *H. influenzae* genomes (for competence) was also analyzed over a range of cluster sizes. We confirmed that the mean cluster silhouettes for the random genes (0.3) were significantly lower than those of the network genes (0.6), $P < 10^{-6}$.

To decide which gene clusters derived from a global application of silhouette were statistically significant and coherent enough to be denoted as evolutionary modules, we used the mean cluster silhouette, described above, as well as D , the mean Euclidean distance between phylogenetic profiles, and C , a coherence measure explained below. We used D to test that the phylogenetic profiles are more similar between module genes than the same number of genes randomly chosen (1,000 iterations) from (i) the genomes as above; or (ii) the network. We also

used C , the coherence measure, to test that module genes approximate an all-or-none distribution, with most species either having nearly all module genes or almost no module genes. P values for significance for all of these measures were calculated by resampling genes repeatedly until the first exception occurred (e.g., until D for the random case exceeded D for the module), and the number of trials recorded. Thus, a gene cluster was denoted an evolutionary module if both its mean cluster silhouette and coherence C were significantly higher and its mean D (Euclidean distance between phylogenetic profiles) significantly lower ($P \leq 10^{-3}$) than for a random gene cluster of the same size (1,000 iterations) drawn from either the *B. subtilis* genome (for sporulation and chemotaxis) or a proportional mixture of the *B. subtilis* and *H. influenzae* genomes (for competence).

Partitioning Sporulation Clusters. Applying clustering and silhouette analysis to the sporulation gene set produced two “modules.” The first module consisted of promiscuous genes like sigma factors, oligopeptide transporters (*opp* and *app* genes), and a few two-component response regulators. The second module was dominated by genes found exclusively in sporulators. Detailed inspection of the gene- and species-clustered phylogenetic profile of the gene set (Fig. S3), and the associated gene dendrogram (Fig. S4) revealed, however, that there are finer structures within the sporulator-specific cluster. These structures were not captured by silhouette analysis, which is a conservative criterion prone to miss fine structure if data sets are structured across multiple levels of resolution. To follow up on this observation, we successively applied the silhouette analysis to the large sporulator-specific cluster, uncovering two tight subclusters and a third diffuse one. Each of these three subclusters had an interesting biological interpretation when examined in the context of the phenotype data we collected. The three clusters discovered thus are the ones reported in Table S5 as well as in the main text of the paper.

The entire “subcluster” analysis above was also applied to chemotaxis and DNA uptake modules; however those modules did not further split into smaller clusters by this method, and hence are reported as such.

Coherence Measure C. Most commonly used distance metrics like Euclidean or Hamming have the feature that they weigh zeros equally to ones when applied to binary data (such as that found in phylogenetic profiles). If the data set has structure on a single level of resolution, or if the data set is not sparse, this is not a problem. If, however, there are multiple levels of resolution, and if one of these “smaller” scales consists of “sparse” vectors with very few ones and many zeros, this can create a problem if the goal is to identify modules. For example, if the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

[1]

were a phylogenetic table representing the presence or absence of genes (rows) in species (columns) and a Euclidean distance metric were applied to the genes, the result would be very short distances between all pairs of genes because most of each vector consists of zeros. These distances would likely be much smaller than one would obtain from a random gene set, and thus a

“significant cluster” in this sense, especially if the gene profiles were rows in a larger non-sparse data set (lots of ones) being analyzed. However, one would not want to call this gene set a module because the genes are not distributed in such a way that most species either have all of the genes or none of them (the fundamental idea of module as evolutionary building block), as in the profile below with the same intergenic distances:

$$A' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}. \quad [2]$$

To avoid the possibility of misclassifying a sparse cluster as a evolutionary module due to this issue, we introduced a “coherence” measure C . C is defined as the percentage of all species with any module genes (over a minimum noise threshold 5%) that have almost all module genes (over 75%). We calculated C for all putative modules, as well as for random gene clusters of the same size (1,000 iterations) drawn from either the *B. subtilis* genome (for sporulation and chemotaxis) or a proportional mixture of the *B. subtilis* and *H. influenzae* genomes (for competence). From these data we calculated P values to test for significant cohesion (putative modules with $P \geq 10^{-3}$ were rejected). Of all of the clusters selected by the silhouette criterion and found otherwise statistically significant and thus candidate modules, one cluster from the competence network was rejected as a module based on failing this “coherence” test (see Fig. S13). The remaining clusters were sufficiently cohesive to be called modules (see Table S5).

Classifying Genes According to Engineering Ontology. Biological stress responses have been previously compared to engineering control systems (9) with separate, well-defined components organized in one or more feedback loops to sense the stress, process information, and regulate and actuate the response. To classify genes as “sensor,” “regulator,” “actuator,” or “cross-talk,” we amalgamated annotation terms from the GO ontology (10), COG functional categories (11), genome annotation, and literature searches on individual pathway genes. We placed genes into a category if they had at least one, but preferably two or more exact matches to terms from that category and no matches with terms from other categories.

Sensor

GO:0007165 “signal transduction” but not “two-component signal transduction”
 GO:0004871 “signal transducer activity”
 GO:0016020 “membrane”
 GO:0016021 “integral to membrane”
 COG0642 “signal transduction histidine kinase”
 COG0840 “methyl-accepting chemotaxis protein”
 COG2202 “FOG: PAS/PAC domain”
 “sensor receptor”

Plus any proteins that control if and when the cell will “exit” from the response (literature search and examination of expression profile time courses).

Regulator

GO:0000160 “two component signal transduction system (phosphorelay)”
 GO:0050906 “detection of stimulus during sensory perception”
 GO:0006352 “transcription initiation”
 GO:0016481 “negative regulation of transcription”
 GO:0016987 “sigma factor activity”
 GO:0006355 “regulation of transcription, DNA-dependent”
 GO:0003700 “transcription factor activity”

GO:0003677 “DNA binding”
 GO:0000166 “nucleotide binding”
 COG functional classes K, D, T
 COG1191 “DNA-directed RNA polymerase specialized sigma subunit”
 COG1192 “ATPases involved in chromosome partitioning”
 COG1989 “DNA-binding protein”
 COG2002 “regulators of stationary/sporulation gene expression”
 COG2201 “chemotaxis response regulator”
 COG0784 “FOG: cheY-like receiver”
 “transcription initiation”
 “transcriptional regulator”

Note that, for classification as a regulator, we do not search on the obvious annotation keyword “DNA-binding” because there are DNA-binding proteins in the competence pathway that bind to single- or double-stranded DNA for the purpose of taking up naked DNA, and not for the purpose of explicitly regulating cellular processes.

Cross-talk

Same classification as regulator, above, plus known to regulate one or more pathways (literature search).

Actuator

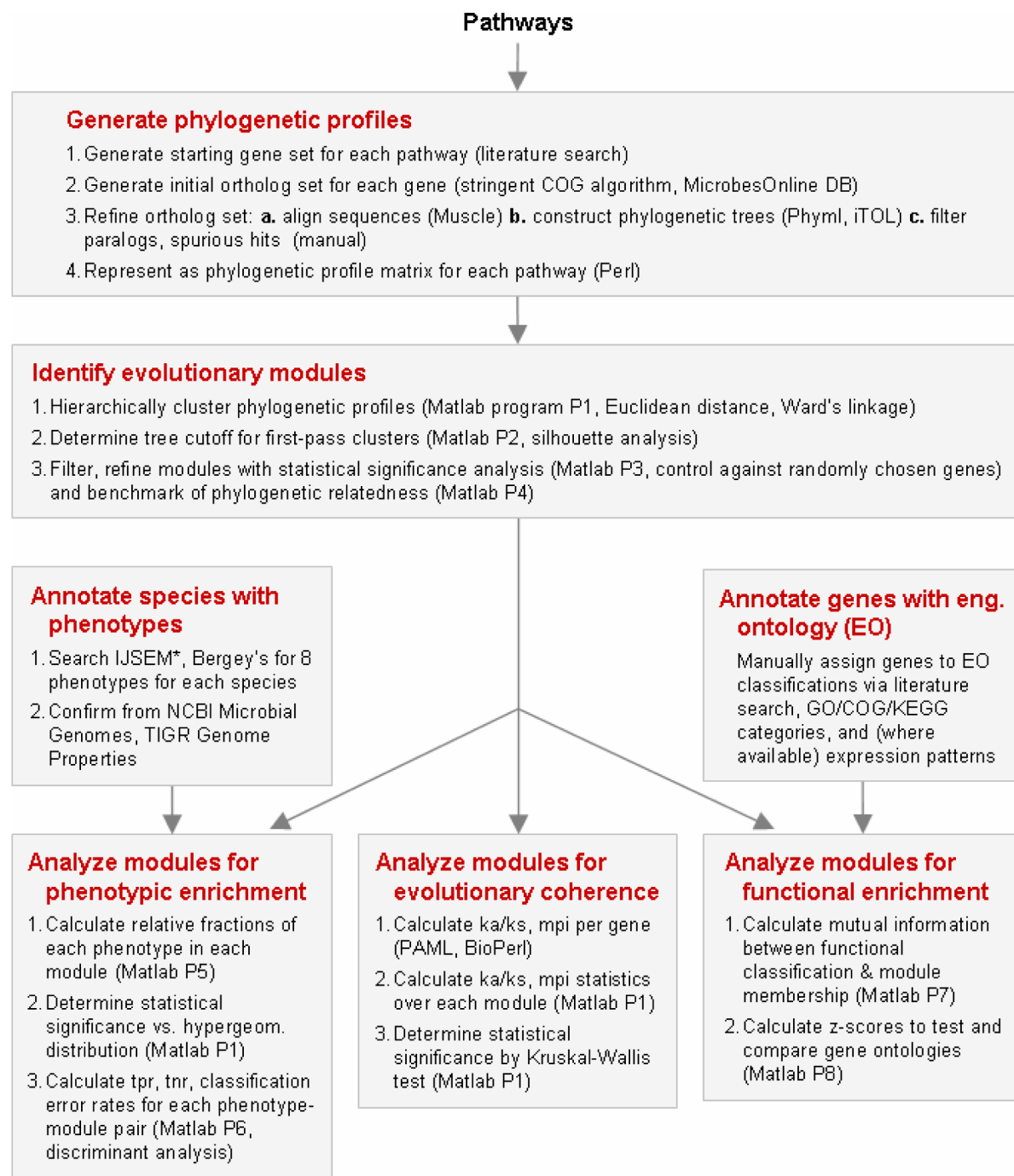
Varies from pathway to pathway because it depends on the specific function of the pathway and what mechanism the cell must employ to respond to the stressor(s). Typically, after all genes in the pathway have been classified as either sensors, regulators, or cross-talk, the actuators are classified as those genes expressed after the main “switch” or commitment signal of the stress response [CheY for chemotaxis (12), Spo0A for sporulation (13), ComK for competence (14)]. For pathways like sporulation, which have multiple switches and checkpoints in the form of sigma factors, the genes under transcriptional control of each sigma factor serve as actuators.

K_a/K_s Calculations. The K_a/K_s ratio for each gene was calculated by an all-pairs algorithm implemented in the Synonymous-Nonsynonymous Alignment Program (15) and also verified by repeating the calculation using the codeml tool of PAML (16) with runmode = -2 and NSSites = 0. Because K_a/K_s tends to saturate for distantly related orthologs that have accumulated multiple synonymous changes after divergence from the last common ancestor (17), we undertook two corrective measures. First, we calculated K_a/K_s values only for gene families that had near-identical phylogenetic profiles, with orthologs in closely related species. This meant that we consistently used the same set of closely related species, which limited any potential variation that might be introduced from distantly related species pairs when averaging pairwise K_a/K_s measurements. For sporulation genes, we calculated K_a/K_s measurements just for orthologs among the 18 spore-formers (Table 1). Second, pairwise gene comparisons for which $K_s > 3$ were discarded before calculating the average K_a/K_s for an ortholog set.

mpi Calculations. To measure the evolutionary coherence of each module, the mean and median percent protein alignment identity (mpi) for each gene was calculated using the Bio package of Perl (www.bioperl.org), and then averaged over all genes in each module (Tables S2–S4). To test whether the mpi for each module was significantly distinct, we used the Kruskal–Wallis nonparametric analysis of variance test (18, Ch. 10). Since the mpi measurements had several outliers (Table S5) and were not necessarily normally distributed, this test was more suitable and robust than one-way ANOVA.

Phenotypic and Functional Enrichment Calculations. P values for functional and phenotypic enrichment were calculated by comparison with the hypergeometric distribution (18).

1. Boone DR, Castenholz RW, Garrity GM, eds (2001) *Bergey's Manual of Systematic Bacteriology* (Springer, New York).
2. Selengut JD, et al. (2007) TIGRFAMs and Genome Properties: Tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* 35:D260–D264.
3. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637.
4. Grebe TW, Stock JB (1999) The histidine protein kinase superfamily. *Adv Microb Physiol* 41:139–227.
5. Aravind L, Walker DR, Koonin EV (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 27:1223–1242.
6. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
7. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
8. Kaufman L, Rousseeuw PJ (2005) *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, Hoboken, NJ).
9. Bourret RB, Stock AM (2002) Molecular information processing: Lessons from bacterial chemotaxis. *J Biol Chem* 277:9625–9628.
10. Ashburner M, et al. (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25–29.
11. Tatusov RL, et al. (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
12. Wadhams GH, Armitage JP (2004) Making sense of it all: Bacterial chemotaxis. *Nat Rev Mol Cell Biol* 5:1024–1037.
13. Sonenshein AL, Hoch JA, Losick R (2002) *Bacillus Subtilis and Its Closest Relatives: From Genes to Cells* (ASM Press, Washington, DC).
14. Dubnau D (1999) DNA uptake in bacteria. *Annu Rev Microbiol* 53:217–244.
15. Korber B (2001) HIV sequence signatures and similarities. *Computational and Evolutionary Analysis of HIV Molecular Sequences*, eds Rodrigo AG, Learn GH (Kluwer, Boston), pp 55–72.
16. Yang Z (1997) PAML: A Program Package for Phylogenetic Analysis by Maximum Likelihood. *Comput Appl Biosci* 13:555–556.
17. Nekrutenko A, Makova KD, Li WH (2002) The K_A/K_S ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. *Genome Res* 12:198–202.
18. Gibbons JD, Chakraborti S (2003) *Nonparametric Statistical Inference* (Marcel Dekker, New York).
19. Moszer I, Jones LM, Moreira S, Fabry C, Danchin A (2002) SubtiList: The reference database for the *Bacillus subtilis* genome. *Nucleic Acids Res* 30:62–65.
20. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280.
21. Alm EJ, et al. (2005) The MicrobesOnline Web site for comparative genomics. *Genome Res* 15:1015–1022.
22. von Mering C, et al. (2005) STRING: Known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33:D433–D437.



Matlab programs: P1: evmodules.m, P2: silh.m, P3: rand_control.m, P4: phylo_benchmark.m, P5: clust_composition.m, P6: sens_spec.m, P7: mut_info.m, P8: zscores.m

*IJSEM: International Journal of Systematic and Evolutionary Microbiology

Fig. S1. Flowchart of methodology and analysis. Code and data are available at <http://genomics.lbl.gov/evmod/>.

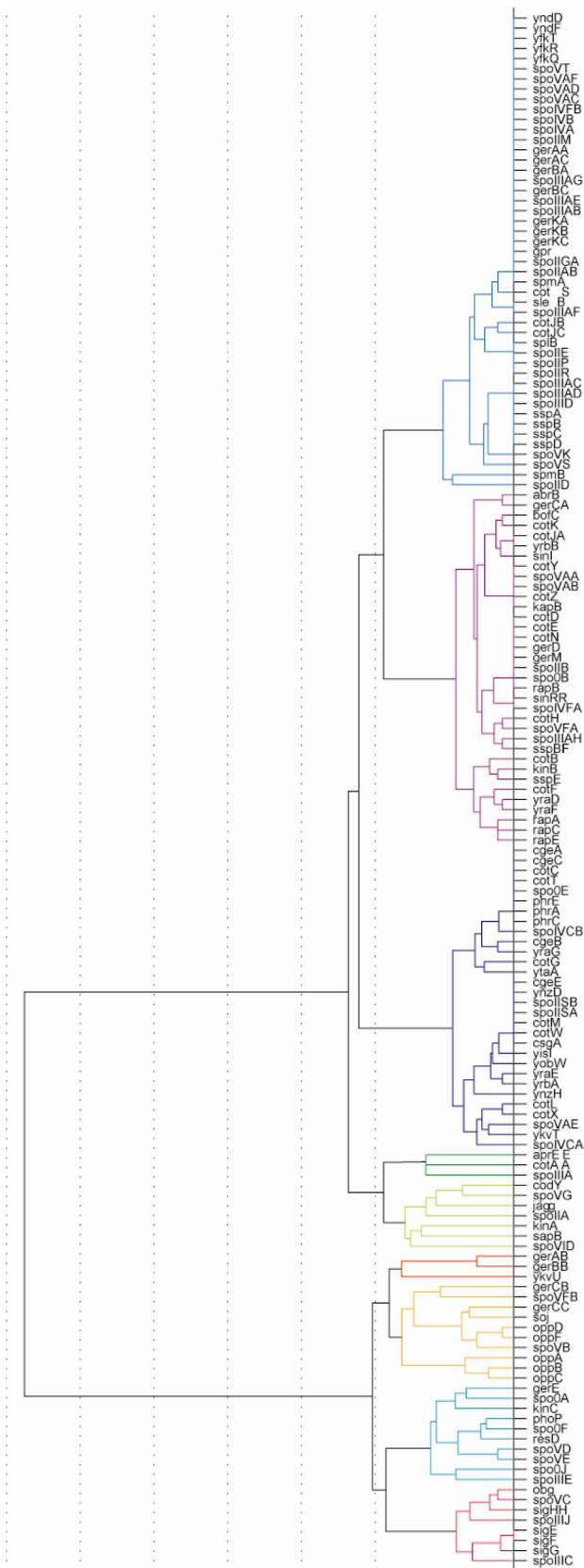


Fig. S4. Sporulation gene dendrogram.

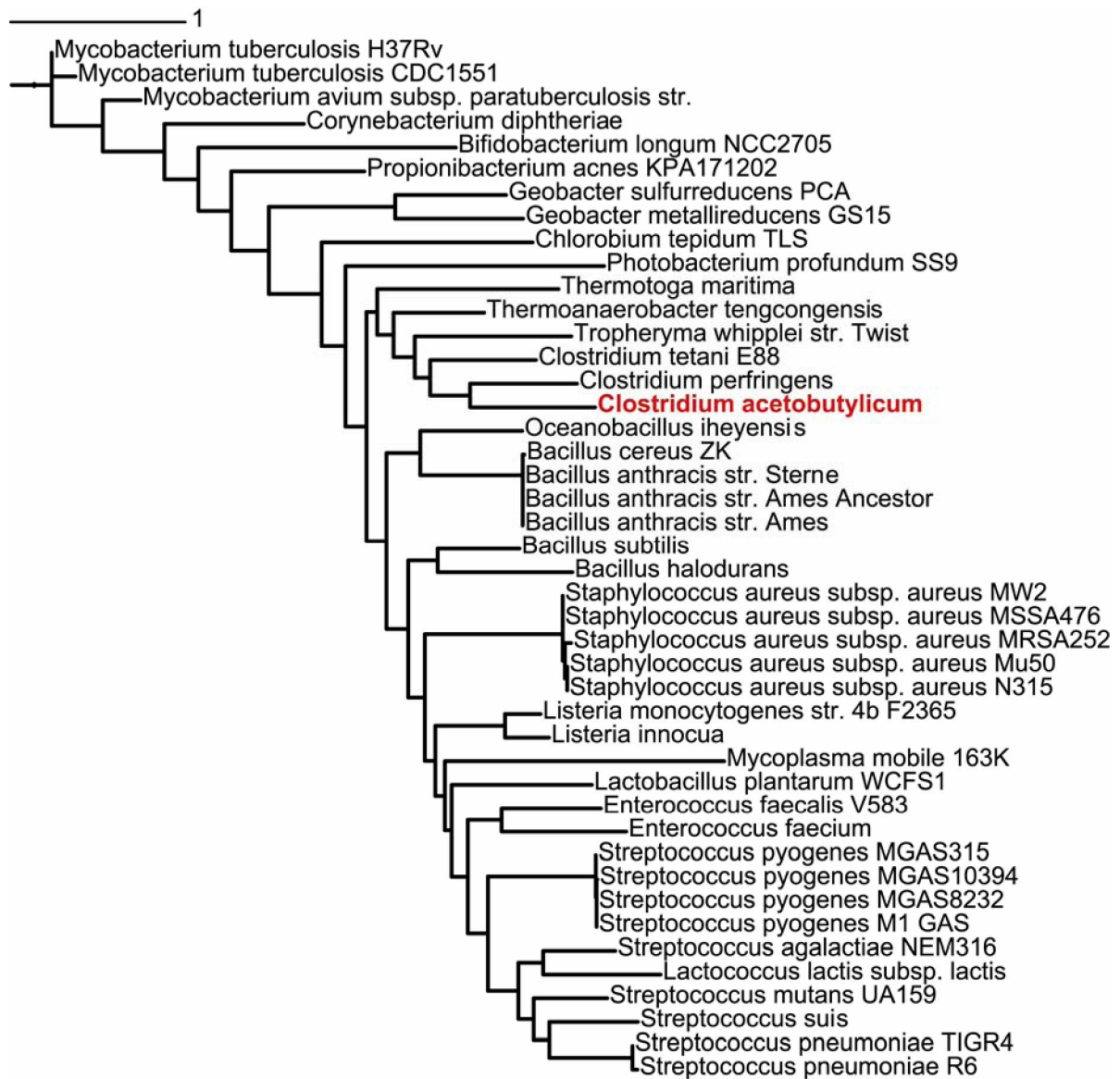


Fig. S9. Maximum-likelihood phylogenetic tree of orthologs of ComEA, an integral membrane protein found in Gram-positive transformable that makes up part of the “pore” which ingests DNA into the cell. *C. acetobutylicum* and related *Clostridia* cluster as expected according to the reference tree.

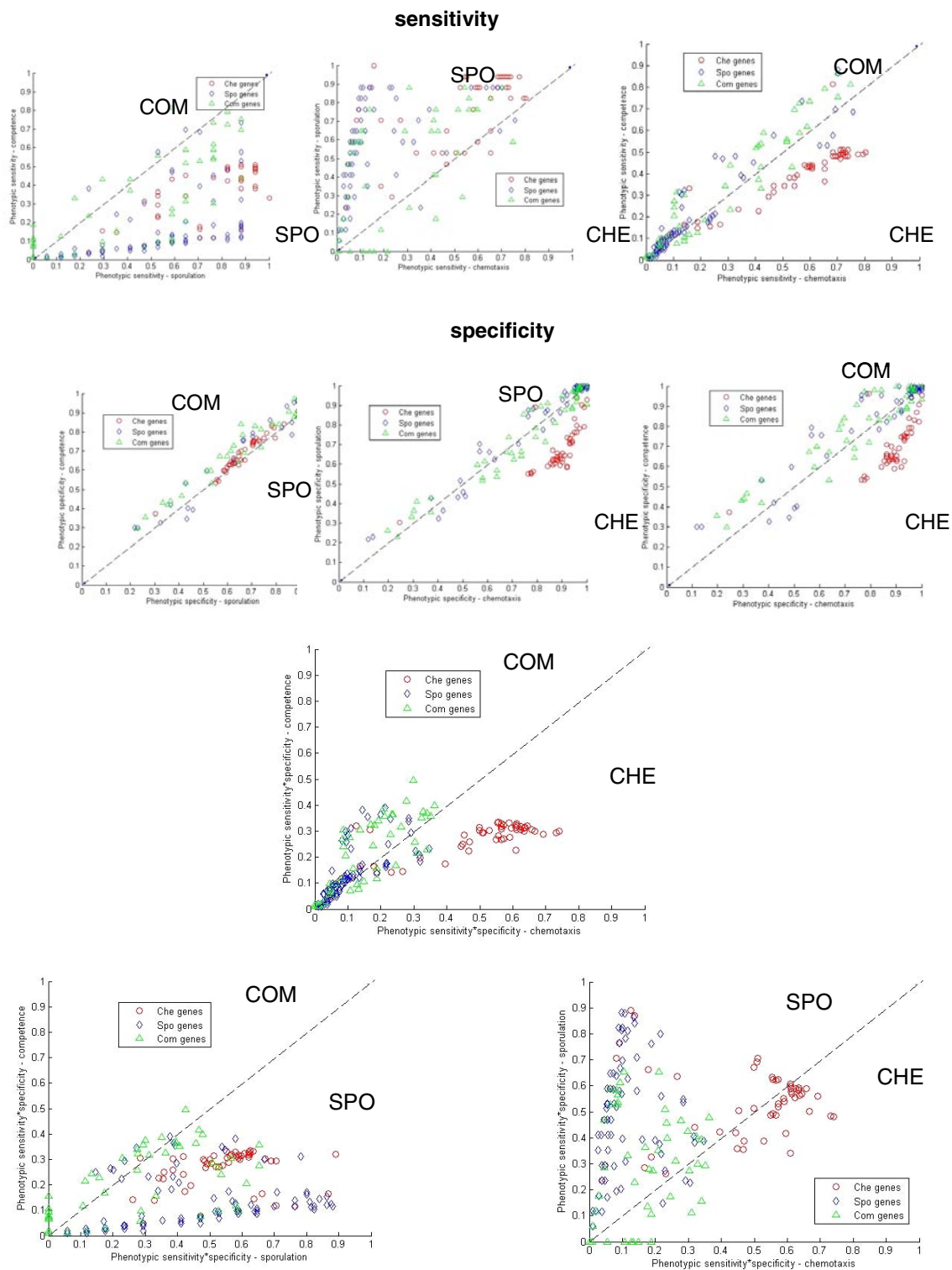


Fig. S12. True positive/true negative rates for all three gene sets and all three cross-phenotype pairs.

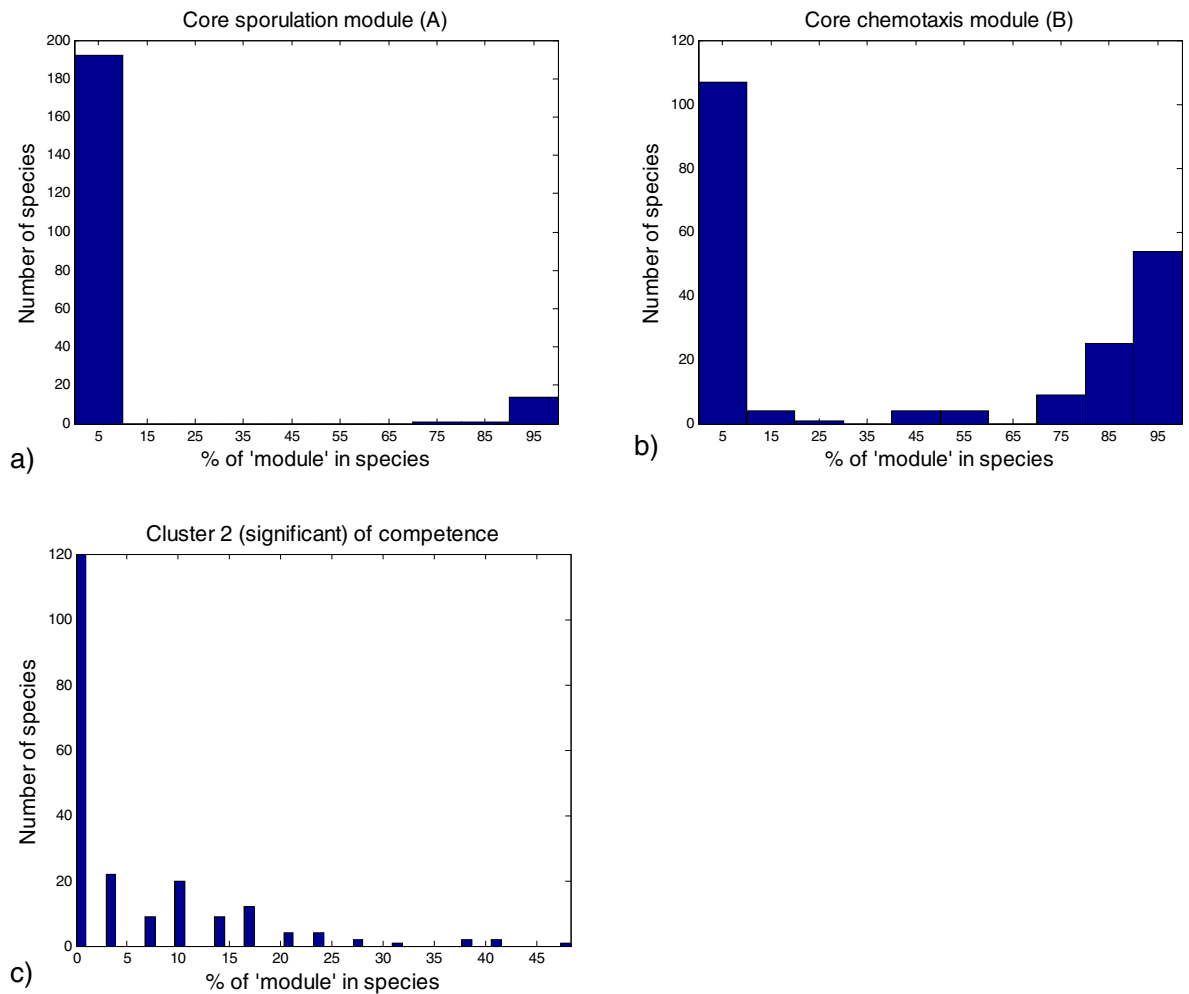


Fig. S13. Cluster coherence histograms. These plots show the distribution of module genes over all bacterial and archaeal species in our data set. Vertical bars show the number of species containing each possible fraction (x axis) of module genes in their genomes. Modules in the chemotaxis (a) and sporulation (b) pathways have the property that most species have either nearly all of the genes in the module or almost none of them, as evidenced by the histogram peaks near 100% and 0% of module genes. This is consistent with the idea of evolutionary modules as evolutionary building blocks a level of resolution above that of genes. Of the two gene clusters within the competence network identified by silhouette analysis, only one was found to be statistically significant (cluster 2). However, we do not call this cluster a evolutionary module because unlike the others, the genes in this set do not have an approximately all-or-none distribution (c). No species has more than about 48% of the cluster genes, and the histogram does not appear bimodal. See [SI Methods](#) and [Table S2](#) for our translation of this “coherence” requirement into the formal metric C.

Table S1. Phenotypes for 207 species. Literature references for each row, as well as a downloadable version of this table, can be found online at <http://www.embl.de/~singh/evmod/>.

No.	Tax ID	Species	C src	Electr. src	Electr. accep.	Habitat	Ani	Pla	Ana	Mot	Com	Spo	Extr	Gm+
1	62977	Acinetobacter sp.ADP1	org	org	O	water, soil, organisms	+	-	-	-	+	-	-	-
2	56636	Aeropyrum pernix	org	inorg	O2	coastal thermal vent	-	-	-	+	-	+	-	-
3	181661	Agrobacterium tumefaciens str.C58 (Cereon)	org	org	O2	soil, plant stems	-	+	-	+	-	-	-	-
4	180835	Agrobacterium tumefaciens str. C58 (U. Washington)	org	org	O2	soil, plant stems	-	+	-	+	-	-	-	-
5	224324	Aquifex aeolicus VF5	inorg	inorg	O2, NO3	deep sea hydrothermal vent	-	-	-	+	-	+	-	-
6	2234	Archaeoglobus fulgidus	inorg	inorg	SO4, S2O3	warm water, subsurface oil fields	-	-	+	+	-	+	-	-
7	354	Azotobacter vinelandii	org		O	soil, water	-	-	-	+	+	+	-	-
8	261594	Bacillus anthracis str. 'Ames Ancestor'				soil, mammalian tissues	+	-	+	-	+	+	-	+
9	198094	Bacillus anthracis str. Ames				soil, mammalian tissues	+	-	+	-	+	+	-	+
10	260799	Bacillus anthracis str. Sterne				soil, mammalian tissues	+	-	+	-	+	+	-	+
11	222523	Bacillus cereus ATCC10987				soil, mammalian tissues	+	-	+	+	+	+	-	+
12	226900	Bacillus cereus ATCC14579				soil, mammalian tissues	+	-	+	+	+	+	-	+
13	288681	Bacillus cereus ZK				soil, mammalian tissues	+	-	+	+	+	+	-	+
14	86665	Bacillus halodurans	org			salt lakes	-	-	+	+	+	+	-	+
15	290609	Bacillus licheniformis ATCC14580	org, inorg	org, inorg		soil, humans	+	-	/	+	+	+	-	+
16	279010	Bacillus licheniformis DSM13	org, inorg	org, inorg		soil, humans	+	-	/	+	+	+	-	+
17	1423	Bacillus subtilis	org	org	O2	soil, water	-	-	+	+	+	+	-	+
18	281309	Bacillus thuringiensis serovar konkukian str. 97-27				soil, insects	+	-	/	+	+	+	-	+
19	817	Bacteroides fragilis	org	org	org	lower intestinal tracts	-	-	+			-	-	-
20	226186	Bacteroides thetaiotaomicron VPI5482	org	org	ferm	mammalian soft tissue	-	-	+			-	-	-
21	283166	Bartonella henselae str. Houston1			O	cats, humans, fleas	+	-	-			-	-	-
22	283165	Bartonella quintana str. Toulouse			O	fixed tissue cells, erythrocytes	+	-	-			-	-	-
23	264462	Bdellovibrio bacteriovorus HD100	org	org	O	soil, sewage, aquatic	-	-	-	+		-	-	-
24	206672	Bifidobacterium longum NCC2705	org	org	ferm	mammalian intestines, feces	-	-	+			-	-	-
25	257310	Bordetella bronchiseptica RB50	org	org	O	animal respiratory tracts	+	-	-	+	+	-	-	-
26	257311	Bordetella parapertussis str12822	org	org	O	human and sheep respiratory tracts	+	-	-	-	+	-	-	-
27	257313	Bordetella pertussis Tohamal	org	org	O	human respiratory tracts	+	-	-	-	+	-	-	-
28	224326	Borrelia burgdorferi B31	org	org	ferm	arthropod, vertebrate hosts	+	-	-	+		-	-	-
29	290434	Borrelia garinii PBI				ticks, animals, humans	+	-	+	+		-	-	-
30	375	Bradyrhizobium japonicum	org	org	O2	soil, plant root hairs	-	-	-	+	+	-	-	-
31	29459	Brucella melitensis	org	org	O2, NO2	hoofed mamals	+	-	-	-		-	-	-
32	204722	Brucella suis str1330	org	org	O2, NO3	swine, reindeer	+	-	-	-		-	-	-
33	107806	Buchnera aphidicola str. APS (Acyrtosiphon pisum)				aphid gut	+	-		-	-	-	-	-
34	224915	Buchnera aphidicola str. Bp (Baizongia pistaciae)				aphid gut	+	-		-	-	-	-	-
35	198804	Buchnera aphidicola str. Sg (Schizaphis graminum)				aphid gut	+	-		-	-	-	-	-
36	243160	Burkholderia mallei ATCC23344				animals, humans	+	-		-	+	-	-	-

No.	Tax ID	Species	C src	Electr. src	Electr. accep.	Habitat	Ani	Pla	Ana	Mot	Com	Spo	Extr	Gm+
37	272560	Burkholderia pseudomallei K96243	org	org	O	soil, animals	+	-	-	+	+	-	-	-
38	266265	Burkholderia xenovorans				animals	+	-		+	+	-	-	-
39	197	Campylobacter jejuni	org	org, inorg		animals	+	-	+	+	+	-	-	-
40	203907	Candidatus Blochmannia floridanus				symbiont of carpenter ants	-	-				-	-	-
41	190650	Caulobacter crescentus CB15	org	org	O2	fresh water, soil	-	-	-	+		-	-	-
42	83560	Chlamydia muridarum	org	org		warm & cold-blooded animals	+	-		-	-	-	-	-
43	813	Chlamydia trachomatis	org	org		warm & cold-blooded animals	+	-		-	-	-	-	-
44	227941	Chlamydomydia caviae GPIC	org	org		warm & cold-blooded animals	+	-			-	-	-	-
45	115711	Chlamydomydia pneumoniae AR39	org	org		warm & cold-blooded animals	+	-				-	-	-
46	115713	Chlamydomydia pneumoniae CWL029					+	-				-	-	-
47	138677	Chlamydomydia pneumoniae J138					+	-				-	-	-
48	182082	Chlamydomydia pneumoniae TW183					+	-				-	-	-
49	194439	Chlorobium tepidum TLS	org	light	HS, S	aquatic environments	-	-	+			-	-	-
50	243365	Chromobacterium violaceum ATCC12472	org	org	org	soil, water, tropics	+	-	/	+		-	-	-
51	1488	Clostridium acetobutylicum	(org)			soil	-	-	+	+		+	-	+
52	1502	Clostridium perfringens	(org)			soil, marine sediments	+	-	+	-		+	-	+
53	212717	Clostridium tetani E88	(org)			human tissues	+	-	+	+		+	-	-
54	1717	Corynebacterium diphtheriae	org	org	O	humans, skin, respiratory tract	+	-	-	-		-	-	+
55	196164	Corynebacterium efficiens YS314	org	org	O2	mammalian skin, mucous membranes	+	-	+			-	-	+
56	196627	Corynebacterium glutamicum ATCC13032	org	org	O2	mammalian skin, mucous membranes	+	-	+			-	-	+
57	227377	Coxiella burnetii RSA493	org	org		ticks, animals	+	-				-	-	-
58	68909	Deinococcus geothermalis				hot spring	-	-	-		+	-	+	+
59	1299	Deinococcus radiodurans	org	org	O2		-	-	+	+	+	-	+	+
60	177439	Desulfotalea psychrophila LSv54	org, inorg	org, h	SO4	marine	-	-	+	+		-	+	
61	207559	Desulfovibrio desulfuricans G20			SO4	soil	-	-	+	+		-	-	-
62	882	Desulfovibrio vulgaris	org, inorg	org, h	inorg, metals	soil	-	-	+	+		-	-	-
63	891	Desulfuromonas spp.				sulfur-rich marine	-	-	+	+		-	-	-
64	226185	Enterococcus faecalis V583	org	org	ferm	soil, mammalian GI tract	+	-	+			-	-	+
65	1352	Enterococcus faecium	org	org	ferm	human, water, soil	+	-	/			-	-	+
66	218491	Erwinia carotovora subsp. atroseptica SCRI1043	org, inorg	org, inorg		plant vascular system, potatoes	-	+	+	+		-	-	-
67	199310	Escherichia coli CFT073	org	org	ferm	GI tract	+	-	+	+		-	-	-
68	83333	Escherichia coli K12	org	org	ferm	GI tract	+	-	+	+		-	-	-
69	155864	Escherichia coli O157:H7 EDL933	org	org	ferm	GI tract	+	-	+	+		-	-	-
70	190304	Fusobacterium nucleatum ATCC25586	org	org	ferm	GI, pulmonary tracts, oral infections	+	-	+			-	-	-
71	28232	Geobacter metallireducens	org	org	feiii	subsurface environments	-	-	+	+		-	-	-
72	243231	Geobacter sulfurreducens PCA	org, inorg	org, inorg	s, org, metals	subsurface environments	-	-	+	+		-	-	-
73	33072	Gloeobacter violaceus	CO2	light	O	rock?	-	-	-			-	-	-
74	233412	Haemophilus ducreyi 35000HP				human mucous membranes	+	-	+		+	-	-	-
75	71421	Haemophilus influenzae Rd KW20	org	org	ferm	mammalian respiratory tract	+	-	+		+	-	-	-

No.	Tax ID	Species	C src	Electr. src	Electr. accep.	Habitat	Ani	Pla	Ana	Mot	Com	Spo	Extr	Gm+
76	272569	Haloarcula marismortui ATCC43049	org	org	O	salt water	-	-	-	+		-	+	-
77	64091	Halobacterium sp. NRC1	org	org	NO3	salt lakes	-	-	+	+		-	+	-
78	235279	Helicobacter hepaticus ATCC51449	org	org	org	mice, GIT, liver	+	-	-	+	+	-	-	-
79	85962	Helicobacter pylori str26695	org	org, inorg	O	mammal, bird GI tracts	+	-	-	+	+	-	-	-
80	85963	Helicobacter pylori J99	org	org, inorg	O	mammal, bird GI tracts	+	-	-	+	+	-	-	-
81	257314	Lactobacillus johnsonii NCC533	org	org	ferm	humans	-	-	/		+	-	-	+
82	220668	Lactobacillus plantarum WCFS1	org	org	ferm	dairy products and environments	-	-	+		+	-	-	+
83	1360	Lactococcus lactis subsp. lactis	org	org	ferm	dairy products and environments	-	-	+	+	+	-	-	+
84	297245	Legionella pneumophila str. Lens			O	soil, freshwater, biofilms	+	-	-	/	+	-	-	-
85	297246	Legionella pneumophila str. Paris			O	soil, freshwater, biofilms	+	-	-	/	+	-	-	-
86	272624	Legionella pneumophila str. Philadelphia1	org	org	ferm	soil, freshwater, biofilms	+	-	-	/	+	-	-	-
87	281090	Leifsonia xyli subsp. xyli str. CTCB07	org	org	org	xylem of sugarcane	-	+	-	-		-	-	+
88	267671	Leptospira interrogans str. Fiocruz L1130	org	org	O2	water, animal renal tissue	+	-	-	+		-	-	-
89	189518	Leptospira interrogans serovar lai str.56601	org	fatty acids		O	-	-		-	+	-	-	-
90	1642	Listeria innocua				soil?	-	-	+	+	+	-	-	+
91	169963	Listeria monocytogenes EGDe	org	org		soil, decaying vegetable matter	+	-	+	+	+	-	-	+
92	265669	Listeria monocytogenes str. 4b F2365	org	org		soil, decaying vegetable matter	+	-	+	+	+	-	-	+
93	221988	Mannheimia succiniciproducens MBEL55E	org	org	org	bovine rumen	-	-	+	-	-	-	-	-
94	265311	Mesoplasma florum L1				mammals, insects, plants	+	+	/	-		-	-	-
95	381	Mesorhizobium loti	org	org	O2	soil, plant root hairs	-	+		+	+	-	-	-
96	2190	Methanocaldococcus jannaschii	CO2	H2	CO2	salt marsh, marine sediments?	-	-	+	+	+	-	+	-
97	39152	Methanococcus maripaludis	CO2	H2	CO2	salt marsh	-	-	+	+	+	-	+	-
98	190192	Methanopyrus kandleri AV19	inorg	inorg H2	S	hydrothermal, volcanic vents	-	-	+	+		-	+	-
99	188937	Methanosarcina acetivorans C2A	inorg	org	CO2	natural waters, anaerobic sewage digestors	-	-	+	+		-	+	-
100	192952	Methanosarcina mazei Goe1	inorg	org	CO2, H2	deep sea hydrothermal vent	-	-	+	+		-	+	-
101	187420	Methanothermobacter thermautotrophicus DeltaH	CO2	H2	CO2	thermophilic,sewage	-	-	+	-		-	+	-
102	243233	Methylococcus capsulatus str. Bath	methane	methane	O	thermophilic	-	-	-	-		-	+	-
103	262316	Mycobacterium avium k10	org	org	O	mammals	+	-	-	-	-	-	-	/
104	233413	Mycobacterium bovis AF212297	org	org	O	humans, animals	+	-	-	-	-	-	-	/
105	1769	Mycobacterium leprae	org	org	O	tissues of warm-blooded animals	+	-	-	-	-	-	-	/
106	83331	Mycobacterium tuberculosis CDC1551	org	org	nitrate?	tissues of warm-blooded animals	+	-	-	-	-	-	-	/
107	83332	Mycobacterium tuberculosis H37Rv	org	org	nitrate?	tissues of warm-blooded animals	+	-	-	-	-	-	-	/
108	233150	Mycoplasma gallisepticum R	org	org		poultry	+	-	/	+	-	-	-	-
109	2097	Mycoplasma genitalium	org	org		vertebrates	+	-	+	-	-	-	-	-
110	295358	Mycoplasma hyopneumoniae str232	org	org		swine	+	-	/	-	-	-	-	-

No.	Tax ID	Species	C src	Electr. src	Electr. accep.	Habitat	Ani	Pla	Ana	Mot	Com	Spo	Extr	Gm+
111	267748	<i>Mycoplasma mobile</i> 163K	org	org	ferm	fish, humans, animals	+	-	/	+	-	-	-	-
112	44101	<i>Mycoplasma mycoides</i> subsp. <i>mycoides</i> SC	org	org	org	cattle	+	-	/	-	-	-	-	-
113	28227	<i>Mycoplasma penetrans</i>	org	org		vertebrates	+	-	+	-	-	-	-	-
114	2104	<i>Mycoplasma pneumoniae</i>	org	org		vertebrates	+	-	+		-	-	-	-
115	2107	<i>Mycoplasma pulmonis</i>	org	org		vertebrates	+	-	+	+	-	-	-	-
116	228908	<i>Nanoarchaeum equitans</i> Kin4M	org	org	org	hyperthermophilic, symbiont with <i>Ignicoccus</i>	-	-	+			-	+	-
117	122586	<i>Neisseria meningitidis</i> MC58	(org)			(cerebrospinal fluid)	+	-	-	-	+	-	-	-
118	122587	<i>Neisseria meningitidis</i> Z2491	(org)			(cerebrospinal fluid)	+	-	-	-	+	-	-	-
119	228410	<i>Nitrosomonas europaea</i> ATCC19718	inorg	inorg	O2	fresh or sea water, soils	-	-	-	+		-	-	-
120	37329	<i>Nocardia farcinica</i>			O	soil, humans, animals	+	-	-	-		-	-	-
121	63737	<i>Nostoc punctiforme</i>	org	org	N?	freshwater	-	-	-	+	+	-	+	-
122	103690	<i>Nostoc</i> sp. PCC7120					-	-	+	+	+	-	+	-
123	182710	<i>Oceanobacillus iheyensis</i>				deep-sea sediment	-	-	-	+	+	+	-	+
124	100379	Onion yellows phytoplasma				phloem of plants	-	+	-			-	-	-
125	264201	<i>Parachlamydia</i> sp. UWE25	org	org		amoebae	-	-	m			-	-	-
126	747	<i>Pasteurella multocida</i>	org	org	O	nasopharynx, gingiva	+	-	+	-	-	-	-	-
127	298386	<i>Photobacterium profundum</i> SS9				moderately deep sea	-	-	/	+	+	-	-	-
128	243265	<i>Photobacterium luminescens</i> subsp. <i>laumondii</i> TTO1	org		ferm	soil nematodes, insects	+	-	+	+		-	-	-
129	263820	<i>Picrophilus torridus</i> DSM9790	org	org	org	thermoacidophilic	-	-	-	-		-	+	-
130	243090	<i>Pirellula</i> sp.1	org	org	org	marine	-	-	-	/	-	-	-	-
131	52972	<i>Polaromonas</i>	inorg	org	O	coal tar contaminated freshwater sediment	-	-	-	-		-	-	-
132	242619	<i>Porphyromonas gingivalis</i> W83	org	org	ferm, org	human oral cavities	+	-	+			-	-	-
133	74547	<i>Prochlorococcus marinus</i> str. MIT9313	org, inorg	light, org	org, inorg	marine	-	-	-	-		-	-	-
134	167539	<i>Prochlorococcus marinus</i> CCMP1375	org, inorg	light	org, inorg	marine	-	-	-	-		-	-	-
135	59919	<i>Prochlorococcus marinus</i> CCMP1986	org, inorg	light	org, inorg	marine	-	-	-	-		-	-	-
136	267747	<i>Propionibacterium acnes</i> KPA171202	org, inorg	org	ferm, inorg	human skin	-	-	m	-		-	-	-
137	208964	<i>Pseudomonas aeruginosa</i> PAO1	org	org	O2	moist or wet environments	+	-		+	+	-	-	-
138	160488	<i>Pseudomonas putida</i> KT2440	org	org	O2	wet soil, plants	-	-	-	+	+	-	-	-
139	223283	<i>Pseudomonas syringae</i> DC3000	org	org	org	wet soil, plants	-	+	-	+	+	-	-	-
140	13773	<i>Pyrobaculum aerophilum</i>	org, inorg	org, inorg	org, inorg	hyperthermophilic	-	-	m	+		-	+	-
141	29292	<i>Pyrococcus abyssi</i>	org	org	inorg s	deep sea hydrothermal vent	-	-	+	+		-	+	-
142	186497	<i>Pyrococcus furiosus</i> DSM3638	org	org	s	deep sea hydrothermal vent	-	-	+	+		-	+	-
143	53953	<i>Pyrococcus horikoshii</i>	org	org	inorg s	deep sea hydrothermal vent	-	-	+	+		-	+	-
144	305	<i>Ralstonia solanacearum</i>	org	org	O2, NO3	water, soil, plants	-	+	-	+	+	-	-	-
145	258594	<i>Rhodopseudomonas palustris</i> CGA009	org, inorg	light, org, inorg	org, inorg	swine waste, earthworm droppings, water	-	-	/	+		-	-	-
146	781	<i>Rickettsia conorii</i>	org	org		arthropod, vertebrate hosts	+	+	+		-	-	-	-
147	782	<i>Rickettsia prowazekii</i>	org	org		arthropod, vertebrate hosts	+	-	+		-	-	-	-

No.	Tax ID	Species	C src	Electr. src	Electr. accep.	Habitat	Ani	Pla	Ana	Mot	Com	Spo	Extr	Gm+
148	257363	Rickettsia typhi str. wilmington				ticks, vertebrates and invertebrates	+	-	-	-		-	-	-
149	90370	Salmonella enterica sv. Typhi	org	org	ferm	GI tract	+	-	+	+	-	-	-	-
150	209261	Salmonella enterica sv. Typhi Ty2	org	org	ferm	human	+	-	+	+	-	-	-	-
151	99287	Salmonella typhimurium LT2	org	org	ferm	GI tract	+	-	+	+	-	-	-	-
152	211586	Shewanella oneidensis MR1			various	soil	-	-	+	+		-	-	-
153	198215	Shigella flexneri 2a str.2457T	org	org	ferm	GI tract	+	-	+	-		-	-	-
154	198214	Shigella flexneri 2a str.301	org				+	-		-		-	-	-
155	292414	Silicibacter TM1040	org	org			-	-	+			-	-	-
156	382	Sinorhizobium meliloti	org	org	O2	soil, plant root hairs	-	+	-	+		-	-	-
157	282458	Staphylococcus aureus subsp. aureus MRSA252				skin of warm-blooded animals	+	-		-	+	-	-	+
158	282459	Staphylococcus aureus subsp. aureus MSSA476				skin of warm-blooded animals	+	-		-	+	-	-	+
159	158878	Staphylococcus aureus subsp. aureus Mu50				skin of warm-blooded animals	+	-		-	+	-	-	+
160	196620	Staphylococcus aureus subsp. aureus MW2				skin of warm-blooded animals	+	-		-	+	-	-	+
161	158879	Staphylococcus aureus subsp. aureus N315				skin of warm-blooded animals	+	-		-	+	-	-	+
162	176280	Staphylococcus epidermidis ATCC12228				skin & mucous membranes of warm-blooded animals	+	-	-	-	+	-	-	+
163	208435	Streptococcus agalactiae 2603VR				human tissues	+	-	+	-	+	-	-	+
164	211110	Streptococcus agalactiae NEM316				human tissues	+	-	+	-	+	-	-	+
165	210007	Streptococcus mutans UA159	org	org	ferm	human tissues	+	-	+	-	+	-	-	+
166	171101	Streptococcus pneumoniae R6	org	org	ferm	human tissues	+	-	+	-	+	-	-	+
167	170187	Streptococcus pneumoniae TIGR4	org	org	ferm	human tissues	+	-	+	-	+	-	-	+
168	160490	Streptococcus pyogenes M1 GAS				human tissues	+	-	+	-	+	-	-	+
169	286636	Streptococcus pyogenes MGAS10394				human tissues	+	-	+	-	+	-	-	+
170	198466	Streptococcus pyogenes MGAS315				human tissues	+	-	+	-	+	-	-	+
171	186103	Streptococcus pyogenes MGAS8232				human tissues	+	-	+	-	+	-	-	+
172	193567	Streptococcus pyogenes SS11				human tissues	+	-	+	-	+	-	-	+
173	286604	Streptococcus suis	org	org	org, N, S	pigs, humans	+	-	/	-	+	-	-	+
174	227882	Streptomyces avermitilis MA4680	org	org	O2	soil, water	-	-	-	-	+	-	-	+
175	2287	Sulfolobus solfataricus	inorg	inorg (S)		acidic mines, hot springs	-	-	+	+		-	+	-
176	111955	Sulfolobus tokodaii	inorg	inorg (S)		acidic mines, hot springs	-	-	+	+		-	+	-
177	292459	Symbiobacterium thermophilum IAM14863	org	org	org, N	compost	-	-	/	+		+	+	-
178	84588	Synechococcus sp. WH8102	org, inorg	org, inorg	org, inorg	marine	-	-	+	+		-	+	-
179	119072	Thermoanaerobacter tengcongensis			ferm	deep sea hydrothermal vent	-	-	+	+	+	+	+	+
180	2303	Thermoplasma acidophilum	org	inorg	O	self-heating coal refuse piles	-	-	-	+	+	-	+	-
181	50339	Thermoplasma volcanium	org	inorg		continental & submarine solfataras	-	-	+	+	+	-	+	-

No.	Tax ID	Species	C src	Electr. src	Electr. accep.	Habitat	Ani	Pla	Ana	Mot	Com	Spo	Extr	Gm+
182	197221	<i>Thermosynechococcus elongatus</i> BP1					-	-	+	+		-	+	-
183	2336	<i>Thermotoga maritima</i>	org	inorg	s2o3	geothermal heated seafloor	-	-	+	+		-	+	-
184	262724	<i>Thermus thermophilus</i> HB27	org	org	org	aquatic thermal	-	-	-	+	+	-	+	/
185	292415	<i>Thiobacillus denitrificans</i>	inorg	inorg	inorg	water and soil	-	-	/	+	+	-	-	-
186	243275	<i>Treponema denticola</i> ATCC35405	org	org	ferm	human oral cavities	+	-	+	+		-	-	-
187	160	<i>Treponema pallidum</i>	org	org	ferm	human	+	-	+	+		-	-	-
188	203267	<i>Tropheryma whipplei</i> str. Twist	org	org	O2	soil, GI tract	+	-	+	-		-	-	-
189	218496	<i>Tropheryma whipplei</i> TW0827					+	-		-		-	-	-
190	2130	<i>Ureaplasma urealyticum</i>	org	org		vertebrates	+	-	+	-		-	-	-
191	666	<i>Vibrio cholerae</i>	org	org	ferm	aquatic	+	-	+	+	+	-	-	-
192	223926	<i>Vibrio parahaemolyticus</i> RIMD2210633	org	org	ferm	marine	+	-	+	+	+	-	-	-
193	216895	<i>Vibrio vulnificus</i> CMCP6	org	org	ferm	marine, wounds	+	-	+	+	+	-	-	-
194	196600	<i>Vibrio vulnificus</i> YJ016	org				+	-	/	+	+	-	-	-
195	36870	<i>Wigglesworthia glossinidia</i>	org	org		arthropod endosymbiont	+	-	+	-	-	-	-	-
196	163164	<i>Wolbachia</i> sp.	org, inorg	org, inorg	org	endosymbiont of <i>Drosophila melanogaster</i>	-	-	-	-	+	-	-	-
197	844	<i>Wolinella succinogenes</i>	org	inorg, org	org	humans, animals	-	-	+	+		-	-	-
198	190486	<i>Xanthomonas axonopodis</i> str.306			O	citrus plants	-	+	-	+	+	-	-	-
199	190485	<i>Xanthomonas campestris</i> ATCC33913	org	org	O2	plant vascular system	-	+	-	+	+	-	-	-
200	160492	<i>Xylella fastidiosa</i> 9a5c				plant xylem, insect	-	+	-	-		-	-	-
201	155920	<i>Xylella fastidiosa</i> ANN1 (oleander)				Oleander bushes	-	+	-	-		-	-	-
202	155919	<i>Xylella fastidiosa</i> DIXON (almond)				Almond trees	-	+	-	-		-	-	-
203	183190	<i>Xylella fastidiosa</i> Temecula1				plant xylem, insect	-	+	-	-		-	-	-
204	229193	<i>Yersinia pestis</i> biovar Medievalis str.91001	org	org	ferm	fleas, mammals	+	-	+	+	+	-	-	-
205	214092	<i>Yersinia pestis</i> CO92	org	org	ferm	fleas, mammals	+	-	+	+	+	-	-	-
206	187410	<i>Yersinia pestis</i> KIM	org	org	ferm	fleas, mammals	+	-	+	+	+	-	-	-
207	273123	<i>Yersinia pseudotuberculosis</i> IP32953	org	org	ferm	animals, humans	+	-	+	+	+	-	-	-
Total							117	17	97	101	85	18	33	47

Table S2. Chemotaxis genes

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	Function	Engineering Ontology
E	1	abrB	Bsu0037	16077105	36570	K	63.3	transcriptional regulator	crosstalk
E	1	codY	Bsu1618	16078680	38152	K	55.4	transcriptional repressor of flagellar gene expression	crosstalk
E	1	fliT	Bsu3529	16080585	40071	-	54.0	flagellar protein; putative chaperone	actuator
E	1	fliY	Bsu1633	16078695	38168	N	26.9	flagellar motor switch protein; CheY~P hydrolyzing protein	actuator
E	1	fliZ	Bsu1635	16078697	38170	N	30.8	required for flagellar formation	unknown
E	1	hpr	Bsu0999	16078063	37532	K	66.2	transcriptional regulator	crosstalk
E	1	sinI	Bsu2458	16079516	38997	-	45.5	antagonist of SinR	crosstalk
E	1	sinR	Bsu2459	16079517	38998	K	68.7	transcriptional regulator	crosstalk
E	1	tgl	Bsu3122	16080179	39663	-	36.7	transglutaminase	unknown
B	2	cheA	Bsu1644	16078706	38179	N	29.9	two-component sensor histidine kinase	enzymatic regulator
B	2	cheB	Bsu1643	16078705	38178	N	39.0	glutamate methylesterase	regulator
B	2	cheR	Bsu2271	16079329	38809	N	27.9	methyltransferase	regulator
B	2	cheW	Bsu1645	16078707	38180	N	23.4	coupling of receptor to kinase	regulator
B	2	cheY	Bsu1634	16078696	38169	T	35.8	two-component response regulator	regulator
B	2	yoaH	Bsu1860	16078921	38397	N	17.8	similar to methyl-accepting chemotaxis protein; possible function in membrane attachment	sensor
B	2	yvaQ	Bsu3366	16080422	39908	N	17.9	similar to transmembrane receptor taxis protein	sensor
-	3	crh	Bsu3471	16080527	40013	G	32.9	catabolite repression HPr-like protein	regulator
-	3	ptsH	Bsu1392	16078454	37925	G	32.2	histidine-containing phosphocarrier protein; PTS system	regulator
-	4	cheC	Bsu1646	16078708	38181	N	20.3	required for correct methylation of McpB; adaptation	regulator
-	4	cheD	Bsu1647	16078709	38182	N	29.6	required for basal activation of CheA kinase	regulator
-	4	cheV	Bsu1403	16078465	37936	T	25.5	enzymatic regulator; promotes kinase activation; adaptation	regulator
-	4	flgM	Bsu3540	16080596	40082	K	23.9	anti-sigma factor repressor of sigma-D-dependent transcription	actuator
-	4	ylxH	Bsu1642	16078704	38177	D	31.0	similar to flagellar biosynthesis switch protein	unknown
A	5	flgB	Bsu1619	16078681	38153	N	23.5	flagellar basal-body rod protein	actuator
A	5	flgC	Bsu1620	16078682	38154	N	37.5	flagellar basal-body rod protein	actuator
A	5	flgE	Bsu1630	16078692	38164	N	45.4	flagellar hook protein	actuator
A	5	flgK	Bsu3538	16080594	40080	N	18.9	flagellar hook-associated protein 1 (HAP1)	actuator
A	5	flgL	Bsu3537	16080593	40079	N	19.9	flagellar hook-associated protein 3 (HAP3)	actuator
A	5	flhA	Bsu1640	16078702	38175	N	36.3	flagella-associated protein	actuator
A	5	flhB	Bsu1639	16078701	38174	N	30.2	flagella-associated protein	actuator
A	5	flhO	Bsu3638	16080693	40180	N	43.8	flagellar basal-body rod protein	actuator
A	5	flhP	Bsu3637	16080692	40179	N	41.2	flagellar hook-basal body protein	actuator
A	5	fliD	Bsu3531	16080587	40073	N	17.7	flagellar hook-associated protein 2 (HAP2)	actuator
A	5	fliE	Bsu1621	16078683	38155	N	24.8	flagellar hook-basal body protein; rod protein	actuator
A	5	fliF	Bsu1622	16078684	38156	N	21.2	flagellar basal-body M-ring protein	actuator
A	5	fliG	Bsu1623	16078685	38157	N	32.0	flagellar motor switch protein; C-ring protein	actuator
A	5	fliI	Bsu1625	16078687	38159	N	44.1	flagellar-specific ATP synthase; delivery complex	actuator
A	5	fliM	Bsu1632	16078694	38167	N	24.1	flagellar motor switch protein; C-ring protein	actuator
A	5	fliP	Bsu1636	16078698	38171	N	41.5	flagellar protein; C-rod protein	actuator
A	5	fliQ	Bsu1637	16078699	38172	N	34.5	flagellar protein; C-rod protein	actuator
A	5	fliR	Bsu1638	16078700	38173	N	22.0	flagellar protein; C-rod protein	actuator
A	5	fliS	Bsu3530	16080586	40072	O	27.3	flagellar protein; putative chaperone	actuator
A	5	hag	Bsu3533	16080589	40075	N	33.9	flagellin protein	actuator
A	5	motA	Bsu1371	16078433	37904	N	27.3	motility protein A; proton conductor	actuator
A	5	motB	Bsu1370	16078432	37903	N	21.2	motility protein B; anchor to the peptidoglycan	actuator
A	5	ylqH	Bsu1609	16078671	38143	S	29.5	similar to flagellar biosynthetic protein	unknown
A	5	ylxG	Bsu1629	16078691	38163	N	21.4	similar to flagellar hook assembly protein	unknown
-	6	flhF	Bsu1641	16078703	38176	N	21.9	GTP-binding signal recognition particle	actuator
-	6	sigD	Bsu1648	16078710	38183	K	33.5	RNA polymerase sigma-28 factor (sigma-D)	transcriptional regulator
-	6	yhfV	Bsu1038	16078102	37571	-	10.3	similar to methyl-accepting chemotaxis protein	sensor
D	7	fliH	Bsu1624	16078686	38158	N	15.3	flagellar assembly protein; delivery complex	actuator
D	7	fliJ	Bsu1626	16078688	38160	N	18.9	flagellar protein; delivery complex chaperone	actuator
D	7	fliK	Bsu1628	16078690	38162	N	11.8	flagellar protein; hook length	actuator
D	7	fliL	Bsu1631	16078693	38166	N	19.9	flagellar protein	actuator
C	8	mcpA	Bsu3119	16080176	39660	N	18.7	glucose, alpha-methyl glucoside receptor	sensor
C	8	mcpB	Bsu3121	16080178	39662	N	18.8	Asn Asp Gln Glu His receptor	sensor

Table S3. Sporulation genes

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
A	1	cotJB	Bsu0690	16077757	37223	-	52.1	0.04	cotJB	actuator
A	1	cotJC	Bsu0691	16077758	37224	P	60.9	0.14	component of spore undercoat	actuator
A	1	cotS	Bsu3085	16080142	39626	-	19.2	0.00	spore coat protein	actuator
A	1	gerAA	Bsu3302	16080358	39844	-	40.7	0.00	germination response to L-alanine and related amino acids (earliest stage)	sensor
A	1	gerAC	Bsu3304	16080360	39846	-	21.6	0.66	germination response to L-alanine and related amino acids (earliest stage)	sensor
A	1	gerBA	Bsu3578	16080633	40120	-	40.8	0.00	germination response to the combination of glucose; fructose; L-asparagine; and KCl (earliest stage)	sensor
A	1	gerBC	Bsu3580	16080635	40122	-	21.7	0.00	germination response to the combination of glucose; fructose; L-asparagine; and KCl (earliest stage)	sensor
A	1	gerKA	Bsu0371	16077438	36904	-	40.7	0.00	germination response to the combination of glucose; fructose; L-asparagine; and KCl	actuator
A	1	gerKB	Bsu0373	16077440	36906	-	17.9	0.54	germination response to the combination of glucose; fructose; L-asparagine; and KCl	actuator
A	1	gerKC	Bsu0372	16077439	36905	-	20.6	0.47	germination response to the combination of glucose; fructose; L-asparagine; and KCl	actuator
A	1	gpr	Bsu2550	16079608	39090	-	59.1	0.19	spore protease	transcriptional regulator
A	1	sapB	Bsu0666	16077733	37199	S	28.0	0.37	mutant activates alkaline phosphatase during sporulation independently of sigma-F and sigma-E	unknown
A	1	sleB	Bsu2292	16079350	38830	M	59.5	0.10	spore cortex-lytic enzyme	enzymatic regulator
A	1	splB	Bsu1395	16078457	37928	L	70.8	0.26	spore photoproduct lyase	enzymatic regulator
A	1	spmA	Bsu2317	16079375	38855	R	53.8	0.17	spore maturation protein	actuator
A	1	spmB	Bsu2316	16079374	38854	S	44.3	0.26	spore maturation protein	actuator
A	1	spolIAA	Bsu2346	16079404	38885	T	22.8	0.31	anti-anti-sigma factor (antagonist of SpoIIAB)	transcriptional regulator
A	1	spolIAB	Bsu2345	16079403	38884	T	63.3	0.19	anti-sigma factor (antagonist of sigma-F) and serine kinase	transcriptional regulator
A	1	spolIID	Bsu3673	16080728	40215	D	25.7	0.33	transcribed in mother cell; mutants block septal thinning	transcriptional regulator
A	1	spolIE	Bsu0064	16077132	36597	T	52.3	0.38	serine phosphatase; required for polar FtsZ ring formation	enzymatic regulator
A	1	spolIGA	Bsu1533	16078595	38066	-	31.4	0.53	protease; converts pro-SigE to SigE	actuator

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
A	1	spolIIAA	Bsu2441	16079499	38980	S	33.6	0.24	mutants block sporulation after engulfment	transcriptional regulator
A	1	spolIIAB	Bsu2440	16079498	38979	-	36.5	0.24	mutants block sporulation after engulfment	actuator
A	1	spolIIAC	Bsu2439	16079497	38978	-	55.2	0.22	mutants block sporulation after engulfment	actuator
A	1	spolIIAD	Bsu2438	16079496	38977	-	48.0	0.30	mutants block sporulation after engulfment	actuator
A	1	spolIIAE	Bsu2437	16079495	38976	-	40.6	0.28	mutants block sporulation after engulfment	actuator
A	1	spolIIAF	Bsu2436	16079494	38975	-	30.9	0.35	mutants block sporulation after engulfment	actuator
A	1	spolIIAG	Bsu2435	16079493	38974	-	35.4	0.42	mutants block sporulation after engulfment	actuator
A	1	spolIID	Bsu3640	16080695	40182	-	73.4	0.09	transcriptional regulator	transcriptional regulator
A	1	spolIIM	Bsu2352	16079410	38891	S	32.5	0.64	mutants do not complete septal thinning	actuator
A	1	spolIP	Bsu2549	16079607	39089	-	40.8	0.39	spolIP	actuator
A	1	spolIR	Bsu3695	16080750	40237	-	45.2	0.39	alternate gene name: ipc-27d csfX	actuator
A	1	spolVA	Bsu2279	16079337	38817	-	67.0	0.16	alternate gene name: spoVP	transcriptional regulator
A	1	spolVB	Bsu2421	16079479	38960	M	48.9	0.22	serine peptidase synthesized in the spore chamber which self-cleaves to regulate sigma _k checkpoint	transcriptional regulator
A	1	spolVFB	Bsu2793	16079849	39333	R	31.4	0.35	membrane metalloprotease governing intercompartmental communication	actuator
A	1	spoVAC	Bsu2341	16079399	38880	-	50.7	0.30	mutants lead to the production of immature spores (stage V sporulation)	actuator
A	1	spoVAD	Bsu2340	16079398	38879	-	48.5	0.24	mutants lead to the production of immature spores (stage V sporulation)	actuator
A	1	spoVAF	Bsu2338	16079396	38877	-	41.3	0.26	mutants lead to the production of immature spores (stage V sporulation)	actuator
A	1	spoVG	Bsu0049	16077117	36582	M	56.2	0.13	required for spore cortex synthesis	actuator
A	1	spoVID	Bsu2807	16079863	39347	-	8.8	0.00	guides safA to spore surface; morphogen	transcriptional regulator
A	1	spoVK	Bsu1743	16078805	38279	O	57.0	0.00	alternate gene name: spoVJ~disruption leads to the production of immature spores (stage V sporulation)	transcriptional regulator

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
A	1	spoVS	Bsu1699	16078761	38234	S	86.9	0.06	required for dehydration of the spore core and assembly of the coat	actuator
A	1	spoVT	Bsu0056	16077124	36589	K	67.7	0.12	transcriptional regulator	transcriptional regulator
A	1	sspA	Bsu2953	16080009	39493	-	69.8	0.23	small acid-soluble spore protein (alpha-type SASP)	actuator
A	1	sspB	Bsu0975	16078040	37508	-	74.6	0.15	small acid-soluble spore protein (beta-type SASP)	actuator
A	1	sspC	Bsu1994	16079053	38532	-	65.7	0.00	small acid-soluble spore protein (alpha/beta-type SASP)	actuator
A	1	sspD	Bsu1349	16078411	37882	-	67.2	0.24	small acid-soluble spore protein (alpha/beta-type SASP)	actuator
A	1	yfkQ	Bsu0779	16077846	37312	-	40.6	0.31	similar to spore germination response	enzymatic regulator
A	1	yfkR	Bsu0778	16077845	37311	-	19.9	0.73	similar to spore germination protein	unknown
A	1	yfkT	Bsu0776	16077843	37309	-	21.1	0.00	similar to spore germination protein	unknown
A	1	yndD	Bsu1776	16078838	38312	-	41.0	0.20	similar to spore germination protein	unknown
A	1	yndF	Bsu1778	16078840	38314	-	22.9	0.51	similar to spore germination protein	unknown
A	1	ytaA	Bsu3087	16080144	39628	-	22.4	0.16	similar to spore coat protein	unknown
B	2	abrB	Bsu0037	16077105	36570	K	74.7	0.10	transcriptional regulator	crossstalk
B	2	bofC	Bsu2771	16079827	39311	-	34.7	0.26	negatively regulates SpoIVB sigK checkpoint	actuator
B	2	cotB	Bsu3603	16080658	40145	-	36.4	0.00	spore coat protein (outer)	actuator
B	2	cotD	Bsu2220	16079278	38758	-	45.2	0.00	spore coat protein (inner)	actuator
B	2	cotE	Bsu1704	16078766	38239	-	58.9	0.23	morphogenic protein	actuator
B	2	cotF	Bsu4050	16081105	40592	-	61.6	0.08	spore coat protein	actuator
B	2	cotH	Bsu3604	16080659	40146	M	46.6	0.00	spore coat protein (inner)	actuator
B	2	cotJA	Bsu0689	16077756	37222	-	77.5	0.00	cotJA	actuator
B	2	cotK	Bsu1800	16078862	38336	-	68.8	0.00	spore coat protein	actuator
B	2	cotN	Bsu2460	16079518	38999	-	41.0	0.24	spore coat-associated protein	actuator
B	2	cotY	Bsu1176	16078240	37709	-	32.7	0.00	spore coat protein (insoluble fraction)	actuator
B	2	cotZ	Bsu1175	16078239	37708	-	33.8	0.23	spore coat protein (insoluble fraction)	actuator
B	2	gerCA	Bsu2275	16079333	38813	-	30.7	0.34	heptaprenyl diphosphate synthase component I	sensor
B	2	gerD	Bsu0155	16077223	36688	-	42.9	0.17	germination response to the combination of glucose; fructose; L-asparagine; and KCl	transcriptional regulator
B	2	gerM	Bsu2834	16079890	39374	R	51.9	0.16	cortex hydrolysis and stage II multiple polar septa; putative role in peptidoglycan synthesis during sporulation	actuator

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
B	2	kapB	Bsu3141	16080198	39682	-	60.6	0.17	kinase-associated protein B	sensor
B	2	kinB	Bsu3140	16080197	39681	T	94.8	0.05	two-component sensor histidine kinase	enzymatic regulator
B	2	rapA	Bsu1244	16078308	37777	R	24.1	0.00	response regulator aspartate phosphatase	enzymatic regulator
B	2	rapB	Bsu3667	16080722	40209	R	26.9	0.00	response regulator aspartate phosphatase	enzymatic regulator
B	2	rapC	Bsu0378	16077445	36911	R	26.2	0.41	response regulator aspartate phosphatase	enzymatic regulator
B	2	rapE	Bsu2578	16079636	39118	R	26.0	0.38	response regulator aspartate phosphatase	enzymatic regulator
B	2	sinI	Bsu2458	16079516	38997	-	45.5	0.00	antagonist of SinR	crossstalk
B	2	sinR	Bsu2459	16079517	38998	K	68.7	0.13	transcriptional regulator	crossstalk
B	2	spo0B	Bsu2789	16079845	39329	-	28.8	0.00	sporulation initiation phosphoprotein	enzymatic regulator
B	2	spoIIB	Bsu2802	16079858	39342	-	18.9	0.48	facilitates septal dissolution; engulfment	actuator
B	2	spoIVFA	Bsu2794	16079850	39334	M	40.7	0.25	inhibitor of SpoIVFB	actuator
B	2	spoVAA	Bsu2343	16079401	38882	-	50.0	0.21	mutants lead to the production of immature spores (stage V sporulation)	actuator
B	2	spoVAB	Bsu2342	16079400	38881	-	60.9	0.18	mutants lead to the production of immature spores (stage V sporulation)	actuator
B	2	spoVFA	Bsu1674	16078736	38209	-	63.6	0.11	dipicolinate synthase subunit A	enzymatic regulator
B	2	sspE	Bsu0865	16077932	37398	-	46.7	0.13	small acid-soluble spore protein (gamma-type SASP)	actuator
B	2	sspF	Bsu0045	16077113	36578	-	71.2	0.10	small acid-soluble spore protein (alpha/beta-type SASP)	actuator
B	2	yraD	Bsu2694	16079752	39234	-	34.3	0.22	similar to spore coat protein	unknown
B	2	yraF	Bsu2691	16079749	39231	-	35.4	0.00	similar to spore coat protein	unknown
B	2	yrbB	Bsu2779	16079835	39319	-	24.6	0.00	similar to spore coat protein	actuator
C	3	cgeA	Bsu1977	16079036	38515	-	100.0	0.00	alternate gene name: cgeAA	actuator
C	3	cgeB	Bsu1978	16079037	38516	S	27.4	0.24	involved in maturation of the outermost layer of the spore	actuator
C	3	cgeC	Bsu1976	16079035	38514	-	100.0	0.00	involved in maturation of the outermost layer of the spore	actuator
C	3	cgeE	Bsu1974	16079033	38512	K	48.4	0.00	involved in maturation of the outermost layer of the spore	transcriptional regulator
C	3	cotC	Bsu1771	16078833	38307	-	100.0	0.00	spore coat protein (outer)	actuator
C	3	cotG	Bsu3605	16080660	40147	-	13.7	0.00	morphogenetic protein	actuator
C	3	cotL	Bsu1799	16078861	38335	-	64.4	0.31	spore coat protein	actuator
C	3	cotM	Bsu1798	16078860	38334	O	51.5	0.00	spore coat protein (outer)	actuator
C	3	cotT	Bsu1210	16078274	37743	-	100.0	0.00	spore coat protein (inner)	actuator
C	3	cotW	Bsu1178	16078242	37711	-	39.0	0.00	spore coat protein (insoluble fraction)	actuator

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
C	3	cotX	Bsu1177	16078241	37710	-	38.2	0.34	spore coat protein (insoluble fraction)	actuator
C	3	csgA	Bsu0208	16077276	36741	-	68.3	0.00	sporulation-specific SASP protein	actuator
C	3	phrA	Bsu1245	16078309	37778	-	100.0	0.00	inhibitor of the activity of phosphatase RapA	sensor
C	3	phrC	Bsu0379	16077446	36912	-	100.0	0.00	regulator of the activity of phosphatase RapC and competence and sporulation stimulating factor (CSF)	sensor
C	3	phrE	Bsu2579	16079637	39119	-	100.0	0.00	regulator of the activity of phosphatase RapE	sensor
C	3	spo0E	Bsu1366	16078428	37899	-	100.0	0.00	negative sporulation regulatory phosphatase	enzymatic regulator
C	3	spolIIAH	Bsu2434	16079492	38973	-	39.1	0.34	mutants block sporulation after engulfment	actuator
C	3	spolISA	Bsu1285	16078348	37818	-	65.7	0.00	alternate gene name: ykaC~lethal when synthesized during vegetative growth in the absence of SpolISB	actuator
C	3	spolISB	Bsu1284	16078347	37817	-	57.1	0.00	disruption blocks sporulation after septum formation	actuator
C	3	spolVCA	Bsu2572	16079630	39112	L	20.6	0.00	site-specific DNA recombinase	enzymatic regulator
C	3	spolVCB	Bsu2571	16079629	39111	K	0.0	0.00	RNA polymerase sporulation-specific sigma factor (sigma-K) (N-terminal half)	transcriptional regulator
C	3	spoVAE	Bsu2339	16079397	38878	-	59.3	0.00	mutants lead to the production of immature spores (stage V sporulation)	actuator
C	3	yisl	Bsu1073	16078137	37606	-	50.0	0.00	phosphatase similar to Spo0E; negative regulator	enzymatic regulator
C	3	ykvT	Bsu1384	16078446	37917	M	30.1	0.32	similar to spore cortex-lytic enzyme	unknown
C	3	ynzD	Bsu1793	16078855	38329	-	63.0	0.00	phosphatase similar to Spo0E; negative regulator	enzymatic regulator
C	3	ynzH	Bsu1768	16078830	38304	-	9.6	0.00	similar to spore coat protein	unknown
C	3	yobW	Bsu1910	16078971	38448	-	33.0	0.00	alternate gene name: csk22~similar to sporulation membrane protein sigma-K-controlled	unknown
C	3	yraE	Bsu2693	16079751	39233	-	41.5	0.00	similar to spore coat protein	unknown
C	3	yraG	Bsu2690	16079748	39230	-	39.9	0.00	similar to spore coat protein	unknown
C	3	yrbA	Bsu2780	16079836	39320	M	50.0	0.00	similar to spore coat protein	unknown
-	4	aprE	Bsu1030	16078094	37563	O	13.0	0.21	serine alkaline protease (subtilisin E)	unknown
-	4	codY	Bsu1618	16078680	38152	K	55.6	0.17	transcriptional regulator	crosstalk
-	4	cotA	Bsu0630	16077697	37163	Q	20.3	0.44	spore coat protein (outer)	enzymatic regulator

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
-	4	gerAB	Bsu3303	16080359	39845	-	16.3	0.00	germination response to L-alanine and related amino acids (earliest stage)	sensor
-	4	gerBB	Bsu3579	16080634	40121	-	15.6	0.38	germination response to the combination of glucose; fructose; L-asparagine; and KCl (earliest stage)	sensor
-	4	gerCB	Bsu2274	16079332	38812	H	31.8	0.10	methyltransferase	sensor
-	4	gerCC	Bsu2273	16079331	38811	H	27.6	0.24	heptaprenyl diphosphate synthase component II	sensor
-	4	gerE	Bsu2837	16079893	39377	K	28.3	0.00	required for the expression of late spore coat genes; transcriptional regulator	transcriptional regulator
-	4	jag	Bsu4100	16081155	40642	R	23.5	0.18	SpolIII-associated protein	transcriptional regulator
-	4	kinA	Bsu1401	16078463	37934	T	15.0	0.00	two-component sensor histidine kinase	enzymatic regulator
-	4	kinC	Bsu1451	16078513	37984	T	19.1	0.00	two-component sensor histidine kinase	enzymatic regulator
-	4	obg	Bsu2788	16079844	39328	R	41.7	0.16	GTPase activity	transcriptional regulator
-	4	oppA	Bsu1144	16078208	37677	E	17.4	0.23	oligopeptide ABC transporter (binding protein)	sensor
-	4	oppB	Bsu1145	16078209	37678	E	31.0	0.30	oligopeptide ABC transporter (permease)	sensor
-	4	oppC	Bsu1146	16078210	37679	E	28.8	0.33	oligopeptide ABC transporter (permease)	sensor
-	4	oppD	Bsu1147	16078211	37680	E	37.5	0.15	oligopeptide ABC transporter (ATP-binding protein)	sensor
-	4	oppF	Bsu1148	16078212	37681	E	35.7	0.00	oligopeptide ABC transporter (ATP-binding protein)	sensor
-	4	phoP	Bsu2907	16079963	39447	T	37.9	0.21	two-component response regulator	transcriptional regulator
-	4	resD	Bsu2311	16079369	38849	T	38.8	0.34	two-component response regulator	transcriptional regulator
-	4	sigE	Bsu1534	16078596	38067	K	36.1	0.10	RNA polymerase sporulation-specific sigma-29 factor (sigma-E); aka SpoIIGB	transcriptional regulator
-	4	sigF	Bsu2344	16079402	38883	K	43.7	0.25	RNA polymerase sporulation-specific sigma factor (sigma-F)	transcriptional regulator
-	4	sigG	Bsu1535	16078597	38068	K	43.5	0.11	RNA polymerase sporulation-specific sigma factor (sigma-G)	transcriptional regulator
-	4	sigH	Bsu0098	16077166	36631	K	19.9	0.11	RNA polymerase sigma-30 factor (sigma-H)	transcriptional regulator
-	4	soj	Bsu4094	16081149	40636	D	27.2	0.15	homologous to ParA; chromosome partitioning	transcriptional regulator
-	4	spo0A	Bsu2420	16079478	38959	T	38.9	0.08	two-component response regulator	transcriptional regulator
-	4	spo0F	Bsu3711	16080766	40253	T	22.9	0.00	two-component response regulator	transcriptional regulator

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	K_a/K_s	Function	Engineering Ontology
-	4	spo0J	Bsu4093	16081148	40635	K	35.5	0.00	antagonist of Soj; homologous to ParB; chromosome partitioning	transcriptional regulator
-	4	spoIIIC	Bsu2634	16079692	39174	K	43.0	0.00	RNA polymerase sporulation-specific sigma factor (sigma-K) (C-terminal half)	transcriptional regulator
-	4	spoIIIE	Bsu1681	16078743	38216	D	31.1	0.21	DNA translocase	transcriptional regulator
-	4	spoIIJ	Bsu4101	16081156	40643	U	20.9	0.16	membrane protein translocase alternate gene name: spo0J87	transcriptional regulator
-	4	spoVB	Bsu2763	16079820	39303	R	12.6	0.31	alternate gene name: spoIIIF	transcriptional regulator
-	4	spoVC	Bsu0053	16077121	36586	J	34.7	0.23	thermosensitive mutant blocks spore coat formation (stage V sporulation)	transcriptional regulator
-	4	spoVD	Bsu1519	16078581	38052	M	23.7	0.52	penicillin-binding protein	transcriptional regulator
-	4	spoVE	Bsu1523	16078585	38056	D	28.0	0.28	spoVE	transcriptional regulator
-	4	spoVFB	Bsu1675	16078737	38210	H	34.1	0.11	dipicolinate synthase subunit B	enzymatic regulator
-	4	ykvU	Bsu1385	16078447	37918	R	0.0	0.00	similar to spore cortex membrane protein	unknown

Table S4. Competence genes

Module No.	Cluster No.	Gene	Gene ID	Genbank ID	VIMSS ID	COG	mpi	Function	Engineering Ontology
-	1	acpD	HI1366	16273276	7706	I	27.4	acyl carrier protein phosphodiesterase	enzymatic regulator
-	1	cinA	Bsu1694	16078756	38229	R	28.6	competence-damage inducible protein	actuator
-	1	comA	Bsu3163	16080219	39704	T	0.0	two-component response regulator	transcriptional regulator
-	1	comC	Bsu2803	16079859	39343	N	0.0	DNA-binding protein	transcriptional regulator
-	1	comE	HI0435	16272383	6785	U	15.5	competence protein E	actuator
-	1	comE1	HI1008	16272943	7358	L	24.0	hypothetical protein	unknown
-	1	comEA	Bsu2555	16079613	39095	L	26.5	integral membrane protein	actuator
-	1	comEB	Bsu2554	16079612	39094	F	21.7	alternate gene name comD	actuator
-	1	comEC	Bsu2553	16079611	39093	R	18.8	putative integral membrane protein	actuator
-	1	comER	Bsu2556	16079614	39096	E	28.5	non-essential gene for competence	actuator
-	1	comF	HI0434	16272382	6784	R	21.0	competence protein F	actuator
-	1	comFA	Bsu3544	16080600	40086	L	21.2	late competence protein	actuator
-	1	comFC	Bsu3542	16080598	40084	R	21.0	late competence gene	actuator
-	1	comGA	Bsu2471	16079529	39010	N	33.8	late competence gene	actuator
-	1	comGB	Bsu2470	16079528	39009	N	24.2	probably part of the DNA transport machinery	actuator
-	1	comM	HI1117	16273042	7463	O	42.5	competence protein	actuator
-	1	comQ	Bsu3166	16080222	39707	H	27.0	transcriptional regulator	transcriptional regulator
-	1	crp	HI0957	16272895	7307	T	15.8	catabolite gene activator	global regulator
-	1	dprA	HI0985	16272923	7337	L	30.9	DNA processing chain A	actuator
-	1	pilA	HI0299	16272254	6655	N	24.1	prepilin peptidase-dependent protein D	actuator
-	1	pilB	HI0298	16272253	6654	N	36.5	protein transport protein	actuator
-	1	pilC	HI0297	16272252	6653	N	24.1	protein transport protein	actuator
-	1	pilD	HI0296	16272251	6652	N	25.4	type 4 prepilin-like protein specific leader peptidase	actuator
-	1	radC	HI0952	30995411	7302	-	33.3	DNA repair protein	actuator
-	1	rec2	HI0061	16272035	6423	R	18.9	recombination protein	actuator
-	1	recA	Bsu1695	16078757	38230	L	58.5	multifunctional SOS repair regulator	actuator
-	1	recF	Bsu0004	16077072	36537	L	14.1	recF	actuator
-	1	recN	Bsu2422	16079480	38961	L	12.2	recN	actuator
-	1	recR	Bsu0021	16077089	36554	L	40.7	alternate gene name recM	actuator
-	1	smf	Bsu1612	16078674	38146	L	30.4	DNA processing Smf protein homolog	actuator
-	1	ssb	HI0250	16272208	6608	L	27.6	single-stranded DNA-binding protein	actuator
-	1	topA	HI1365	16273275	7705	L	22.6	DNA topoisomerase I	actuator
-	1	-	HI0365	16272313	6715	R	36.2	hypothetical protein	unknown
-	2	comA	HI0439	16272387	6789	-	32.6	competence protein A	actuator
-	2	comB	HI0438	16272386	6788	-	25.0	competence protein B	actuator
-	2	comC	HI0437	16272385	6787	-	22.5	competence protein C	actuator
-	2	comD	HI0436	16272384	6786	-	100.0	competence protein D	actuator
-	2	comFB	Bsu3543	16080599	40085	-	62.8	late competence gene	actuator
-	2	comGC	Bsu2469	16079527	39008	U	25.5	comGC	actuator
-	2	comGD	Bsu2468	16079526	39007	N	22.5	probably part of the DNA transport machinery	actuator
-	2	comGE	Bsu2467	16079525	39006	-	28.7	probably part of the DNA transport machinery	actuator
-	2	comGF	Bsu2466	16079524	39005	U	22.9	probably part of the DNA transport machinery	actuator
-	2	comGG	Bsu2465	16079523	39004	-	34.7	probably part of the DNA transport machinery	actuator
-	2	comK	Bsu1042	16078106	37575	K	31.6	competence transcription factor	crossstalk
-	2	comP	Bsu3164	16080220	39705	T	25.6	two-component sensor histidine kinase	unknown
-	2	comS	Bsu0351	16077419	36884	-	100.0	comS	transcriptional regulator
-	2	comX	Bsu3165	16080221	39706	-	100.0	competence pheromone precursor (aa 46–55)	sensor
-	2	cyaA	HI0604	16272546	6951	F	44.3	adenylate cyclase	enzymatic regulator
-	2	mecA	Bsu1153	16078217	37686	O	25.4	mecA	transcriptional regulator
-	2	nin	Bsu0343	16077411	36876	-	47.2	alternate gene name comJ	actuator
-	2	nucA	Bsu0344	16077412	36877	-	96.5	nuclease	actuator
-	2	phrC	Bsu0379	16077446	36912	-	100.0	regulator of RapC phosphatase activity; CSF	sensor
-	2	tfoX	HI0601	16272544	6947	K	30.3	DNA transformation protein	actuator
-	2	yphH	Bsu2296	16079354	38834	O	27.7	similar to negative regulation of competence MecA	transcriptional regulator
-	2	yuxH	Bsu3169	16080225	39710	T	28.7	alternate gene name comB; yufA	actuator
-	2	-	HI0938	16272875	7287	N	38.6	hypothetical protein	enzymatic regulator
-	2	-	HI0939	16272876	7288	U	25.6	hypothetical protein	unknown
-	2	-	HI0940	16272877	7289	-	33.8	hypothetical protein	unknown
-	2	-	HI0941	16272878	7290	-	45.3	hypothetical protein	unknown
-	2	-	HI1182	16273520	7527	-	35.2	hypothetical protein	actuator
-	2	-	HI1183	16273105	7528	-	35.8	hypothetical protein	actuator
-	2	-	HI1631	16273520	7962	-	100.0	hypothetical protein	unknown

Table S5. Statistics for chemotaxis, sporulation, and competence modules. Abbreviations in the column headings: D represents the modularity of a gene group, calculated as the mean of Euclidean distances between each pair of gene phylogenetic profiles. C represents the module coherence, defined as the percent of all species with any module genes (over the minimum noise threshold 5–10%) that have almost all module genes (over 75%). P_g is the probability of functional enrichment as compared to the hypergeometric distribution. mpi represents the level of sequence conservation, calculated as the median protein identity. K_a/K_s is the ratio of synonymous to nonsynonymous nucleotide substitutions. Standard deviations are reported in parentheses. Clusters are shown in bold if they were not statistically significant or coherent enough to be called modules, because of either low silhouette, or high D or low coherence C resulting in a P value greater than 10^{-3}

Module No.	Cluster No.	No. of Genes	Mean silhouette	D, Intra-module distance		D Random, Genome	P(D lower than random)	D Random, Network	P(D lower than network)	C, % module coherence	C, random	P(C greater than random)	Most frequent function	P_g	mpi
				Mean silhouette	Intra-module distance										
E	1	9	0.47	3.41 (1.36)	8.65 (0.99)	5.29e-04	1.02e-04	6.45 (0.70)	1.02e-04	37.0	5.9	7.35e-05	Crosstalk	1.51e-06	52.1
B	2	7	0.26	4.07 (1.05)	8.67 (0.95)	4.39e-03	1.03e-14	6.33 (0.59)	1.03e-14	78.2	4.3	<1.00e-05	Regulator	7.12e-03	27.4
	3	2	1.00	-	-	-	-	-	-	-	-	-	Crosstalk	5.74e-02	32.6
A	4	5	-0.01	6.01 (0.83)	8.80 (1.04)	1.79e-02	<1.00e-05	6.47 (0.41)	<1.00e-05	85.3	1.6	<1.00e-05	Regulator	5.05e-04	42.5
	5	24	0.38	3.43 (1.05)	8.69 (0.54)	1.79e-02	<1.00e-05	6.47 (0.41)	<1.00e-05	-	-	-	Actuator	9.99e-12	29.7
D	6	3	-0.10	6.47 (0.42)	9.17 (1.41)	3.33e-01	3.33e-01	-	3.33e-01	-	-	-	Regulator	4.31e-01	40.7
	7	4	0.21	4.80 (0.59)	8.98 (1.48)	4.00e-03	3.33e-02	6.90 (1.04)	3.33e-02	43.6	4.5	6.90e-05	Actuator	5.25e-02	16.7
C	8	7	0.43	3.19 (0.99)	8.75 (1.08)	1.86e-03	6.41e-04	6.48 (0.80)	6.41e-04	74.7	4.3	<1.00e-05	Sensor	2.75e-07	18.6
	Total	61	0.33	6.58 (2.13)	8.69 (0.33)	<1.00e-05	-	-	-	-	-	-	-	-	30.0
Module No.	Cluster No.	No. of Genes	Mean silhouette	D, Intra-module distance		D Random, Genome	P(D lower than random)	D Random, Network	P(D lower than network)	C, module coherence	C, random	P(C greater than random)	Most frequent function	P_g	K_a/K_s
		Mean silhouette	Intra-module distance												
A	1	47	0.78	1.12 (1.21)	8.69 (0.37)	<1.00e-05	<1.00e-05	6.38 (0.77)	<1.00e-05	100.0	1.3	<1.00e-05	Actuator	2.15e-02	54.6
	2	35	0.78	1.70 (0.70)	8.64 (0.49)	<1.00e-05	<1.00e-05	7.07 (0.36)	<1.00e-05	75.0	1.7	<1.00e-05	Actuator	4.50e-02	58.6
C	3	30	0.74	1.90 (0.92)	8.67 (0.45)	<1.00e-05	1.16e-05	6.37 (0.97)	1.16e-05	16.7	1.8	<1.00e-05	Actuator	7.93e-02	28.5
	4	31	0.46	7.99 (2.71)	8.86 (0.31)	7.69e-02	-	5.79 (0.53)	-	-	-	-	Regulator	2.49e-06	49.8
Total	-	153	0.69	6.42 (4.73)	8.69 (0.20)	<1.00e-05	-	-	-	-	-	-	-	49.1	0.24
Module No.	Cluster No.	No. of Genes	Mean silhouette	D, Intra-module distance		D Random, Genome	P(D lower than random)	D Random, Network	P(D lower than network)	C, module coherence	C, random	P(C greater than random)	Most frequent function	P_g	mpi
		Mean silhouette	Intra-module distance												
1	1	33	0.28	7.79 (1.64)	8.79 (0.38)	5.00e-02	1.11e-01	8.39 (0.37)	1.11e-01	52.0	1.4	<1.00e-05	Actuator	3.15e-02	25.2
	2	29	0.64	3.96 (1.81)	8.66 (0.49)	<1.00e-05	<1.00e-05	8.47 (0.37)	<1.00e-05	0.0	1.9	inf	Actuator	3.15e-02	46.5
Total	-	62	0.45	8.61 (3.24)	9.08 (0.12)	1.11e-01	-	-	-	-	-	-	-	35.2	-

Table S6. Z-scores of mutual information between ontology and clustering. The engineering ontology agrees best with the clustering ($z = 16.14$) while the COG Number agrees least ($z = 1.01$). Additional hierarchical ontologies (GO (10), Subtilist (19), KEGG (20)) were also used (data not shown), with qualitatively similar results.

Gene function given by:	Che	Spo	Com
Engineering Ontology	16.14	10.08	0.64
TIGR Primary Role Category	4.38	-	5.45
TIGR Secondary Role Category	5.17	1.18	5.49
COG Number	1.01	0.64	6.05
COG Functional Class	6.36	7.42	6.49

Table S7. Percent classification error of the sensitivity and specificity of genes in each network against each phenotype, using linear discriminant analysis. In general, individual modules in each network are more sensitive and specific to their own phenotype than they are to other phenotypes. However, chemotaxis module E classifies better with the sporulation phenotype (7%) than the motility phenotype (40%). We tested classification error for module genes against combinations of phenotypes (motility AND sporulation, sporulation OR competence, and so on (data not shown). We also tested classification error for combinations of modules (chemotaxis module A AND B, B NOT D, and so on) against individual phenotypes (data not shown). For additional classification results, please see [Fig. S12](#)

Genes	Percent classification error against phenotype		
	Mot	Spo	Com
Chemotaxis	8	11	17
Module A	17	28	28
Module B	12	29	39
Module C	18	19	34
Module D	26	33	41
Module E	40	7	36
Sporulation	5	0	1
Module A	28	6	27
Module B	31	6	34
Module C	30	9	35
Competence	14	9	13
Cluster 2	18	13	20

Table S8. Genes in *Clostridium acetobutylicum*, a non-transformable bacterium, that are homologous to competence genes. The mean %GC of the genome is 0.32, which is not significantly different from the %GC of any of these genes, indicating that either the genes were part of a larger repertoire of once-functional competence genes, or they were laterally transferred so long ago that their %GC has ameliorated to baseline, genomic levels. Please see Figs. S12–S13 for phylogenetic trees for each of these genes.

Genbank ID	Gene	%GC	Most Homologous To . . .	
15896820	cinA	0.36	<i>B. subtilis</i>	62%
15894550	comEA	0.31	<i>S. pneumoniae</i>	57%
15896039	comC	0.28	<i>B. subtilis</i>	42%
			<i>L. innocua</i>	41%
15895042	comM	0.32	<i>H. influenzae</i>	60%

Table S9. Additional genes with similar phylogenetic profiles to chemotaxis, sporulation, and competence genes.

	Genbank ID	Gene Name	Gene ID	COG Fn.	COG No.	Description
Chemotaxis						
Module A	16079176	yomZ	BSU2117	-	-	hypothetical protein
	16080025	ytxD	BSU2973	N	COG1291	hypothetical protein
	16080474	yvfG	BSU3421	-	-	hypothetical protein
	16080595	yvyG	BSU3542	-	-	hypothetical protein
	16081088	yyxA	BSU4036	O	COG0265	hypothetical protein
Chemotaxis						
Module B	16077804	yfmR	BSU0737	R	COG0488	hypothetical protein
	16078460	ykwC	BSU1396	I	COG2084	hypothetical protein
	16079176	yomZ	BSU2117	-	-	hypothetical protein
	16080025	ytxD	BSU2973	N	COG1291	hypothetical protein
	16080595	yvyG	BSU3542	-	-	hypothetical protein
Chemotaxis						
Module C	16077804	yfmR	BSU0737	R	COG0488	hypothetical protein
	16078460	ykwC	BSU1396	I	COG2084	hypothetical protein
	16079176	yomZ	BSU2117	-	-	hypothetical protein
	16080025	ytxD	BSU2973	N	COG1291	hypothetical protein
Chemotaxis						
Module D	16079176	yomZ	BSU2117	-	-	hypothetical protein
	16080025	ytxD	BSU2973	N	COG1291	hypothetical protein
	16080474	yvfG	BSU3421	-	-	hypothetical protein
	16080595	yvyG	BSU3542	-	-	hypothetical protein
Chemotaxis						
Module E	16077102	yabB	BSU0034	R	COG4123	hypothetical protein
	16077116	yabJ	BSU0048	J	COG0251	hypothetical protein
	16077125	yabM	BSU0057	R	COG2244	hypothetical protein
	16077131	yabR	BSU0063	J	COG1098	hypothetical protein
	16077164	yacO	BSU0096	J	COG0566	hypothetical protein
	16077245	ybbT	BSU0177	G	COG1109	hypothetical protein
	16077394	ycgR	BSU0325	R	COG0701	hypothetical protein
	16077920	yfhH	BSU0853	-	-	hypothetical protein
	16077855	yfkJ	BSU0788	T	COG0394	hypothetical protein
	16077846	yfkQ	BSU0779	-	-	hypothetical protein
	16077935	ygaD	BSU0869	V	COG1132	hypothetical protein
	16078046	yhaZ	BSU0981	L	COG4335	hypothetical protein
	16078166	yitK	BSU1102	S	COG1666	hypothetical protein
	16078218	yjbF	BSU1153	R	COG4469	hypothetical protein
	16078224	yjbL	BSU1159	-	-	hypothetical protein
	16078519	ykrA	BSU1455	R	COG0561	hypothetical protein
	16078529	yktB	BSU1465	S	COG4493	hypothetical protein
	16078475	ykuK	BSU1411	S	COG1978	hypothetical protein
	16078530	ykzI	BSU1466	-	-	hypothetical protein
	16078561	ylbD	BSU1497	-	-	hypothetical protein
	16078564	ylbG	BSU1500	S	COG4471	hypothetical protein
	16078565	ylbH	BSU1501	L	COG0742	hypothetical protein
	16078589	ylxW	BSU1525	S	COG3879	hypothetical protein
	16078839	yndE	BSU1776	-	-	hypothetical protein
	16078841	yndG	BSU1778	-	-	hypothetical protein
	16078859	yneK	BSU1796	-	-	hypothetical protein
	16078868	yneS	BSU1807	S	COG0344	hypothetical protein
	16079176	yomZ	BSU2117	-	-	hypothetical protein
	16079121	yoqI	BSU2062	-	-	hypothetical protein
	16078913	yoxB	BSU1852	-	-	hypothetical protein
	16079355	yphG	BSU2298	R	COG1408	hypothetical protein
	16079338	yphF	BSU2281	-	-	hypothetical protein
	16079316	yphA	BSU2259	R	COG0457	hypothetical protein
	16079286	yppD	BSU2228	-	-	hypothetical protein
	16079277	yphA	BSU2219	S	COG4474	hypothetical protein
	16079363	ypzE	BSU2306	-	-	hypothetical protein
	16079590	yqfC	BSU2536	-	-	hypothetical protein
	16079543	yqgP	BSU2487	R	COG0705	hypothetical protein
	16079452	yqiZ	BSU2396	E	COG1126	hypothetical protein
	16079793	yrrK	BSU2739	L	COG0816	hypothetical protein
	16080116	ytkC	BSU3064	R	COG4824	hypothetical protein

	Genbank ID	Gene Name	Gene ID	COG Fn.	COG No.	Description
	16079980	ytnM	BSU2928	R	COG0730	hypothetical protein
	16080036	ytpP	BSU2984	O	COG0526	hypothetical protein
	16080090	ytsC	BSU3038	V	COG1136	hypothetical protein
	16079949	ytxB	BSU2897	S	COG0398	hypothetical protein
	16080523	yvcR	BSU3470	V	COG1136	hypothetical protein
	16080576	yvzD	BSU3523	-	-	hypothetical protein
	16080817	ywfkK	BSU3765	K	COG0583	hypothetical protein
	16080745	ywlF	BSU3692	G	COG0698	hypothetical protein
	16080688	ywpD	BSU3635	T	COG0784	hypothetical protein
	16080636	ywtG	BSU3583	G	COG0477	hypothetical protein
	16081015	yxdL	BSU3964	V	COG1136	hypothetical protein
	16081104	yybS	BSU4052	S	COG4241	hypothetical protein
	16081091	yycH	BSU4039	S	COG4863	hypothetical protein
Sporulation Module A	16077085	yaal	BSU00170	Q	COG1335	hypothetical protein
	16077225	ybaN	BSU01570	G	COG0726	hypothetical protein
	16077497	ydaM	BSU04300	M	COG1215	hypothetical protein
	16077603	ydfC	BSU05360	G	COG0697	hypothetical protein
	16077639	ydhE	BSU05720	G	COG1819	hypothetical protein
	16077792	yetO	BSU07250	P	COG0369	hypothetical protein
	16077879	yfjF	BSU08120	S	COG1742	hypothetical protein
	16077861	yfkC	BSU07940	M	COG0668	hypothetical protein
	16077834	yflI	BSU07670	-	-	hypothetical protein
	16077806	yfmP	BSU07390	K	COG0789	hypothetical protein
	16078064	yhaH	BSU10000	R	COG4980	hypothetical protein
	16078044	yheB	BSU09790	S	COG4399	hypothetical protein
	16078043	yheC	BSU09780	-	-	hypothetical protein
	16078194	yjaV	BSU11290	-	-	hypothetical protein
	16078245	yjcB	BSU11800	-	-	hypothetical protein
	16078247	yjcD	BSU11820	L	COG0210	hypothetical protein
	16078292	yjlB	BSU12270	S	COG4297	hypothetical protein
	16078197	yjzB	BSU11320	-	-	hypothetical protein
	16078192	yjzD	BSU11270	-	-	hypothetical protein
	16078531	yktC	BSU14670	G	COG0483	hypothetical protein
	16078477	ykuL	BSU14130	R	COG0517	hypothetical protein
	16078481	ykuP	BSU14170	C	COG0716	hypothetical protein
	16078541	ylaG	BSU14770	T	COG1217	hypothetical protein
	16078543	ylal	BSU14790	S	COG4896	hypothetical protein
	16078559	ylbB	BSU14950	R	COG0517	hypothetical protein
	16078562	ylbE	BSU14980	-	-	hypothetical protein
	16078563	ylbF	BSU14990	S	COG3679	hypothetical protein
	16078574	ylbP	BSU15100	K	COG0454	hypothetical protein
	16078854	yneF	BSU17910	S	COG3763	hypothetical protein
	16078924	yoaK	BSU18640	S	COG3619	hypothetical protein
	16078994	yocR	BSU19340	R	COG0733	hypothetical protein
	16079020	yodJ	BSU19620	M	COG1876	hypothetical protein
	16079176	yomZ	BSU21170	-	-	hypothetical protein
	16079026	yoze	BSU19680	S	COG4479	hypothetical protein
	16079261	ypbR	BSU22030	R	COG0699	hypothetical protein
	16079255	ypeQ	BSU21960	-	-	hypothetical protein
	16079344	ypgA	BSU22870	C	COG1304	hypothetical protein
	16079339	yphE	BSU22820	-	-	hypothetical protein
	16079310	ypjA	BSU22530	S	COG4347	hypothetical protein
	16079285	yppE	BSU22270	-	-	hypothetical protein
	16079284	yppF	BSU22260	-	-	hypothetical protein
	16079279	yprB	BSU22210	L	COG3359	hypothetical protein
	16079378	yphu	BSU23210	K	COG1386	hypothetical protein
	16079568	yqfS	BSU25130	L	COG0648	hypothetical protein
	16079564	yqfW	BSU25090	-	-	hypothetical protein
	16079562	yqfY	BSU25070	I	COG0821	hypothetical protein
	16079537	yqgV	BSU24810	S	COG0011	hypothetical protein
	16079535	yqgX	BSU24790	R	COG0491	hypothetical protein
	16079507	yqhO	BSU24510	R	COG1752	hypothetical protein
	16079505	yqhQ	BSU24490	R	COG3872	hypothetical protein
	16079468	yqiQ	BSU24120	G	COG2513	hypothetical protein

	Genbank ID	Gene Name	Gene ID	COG Fn.	COG No.	Description
	16079784	yrrS	BSU27300	-	-	hypothetical protein
	16080061	yteU	BSU30090	-	-	hypothetical protein
	16080044	ytmP	BSU29920	M	COG0510	hypothetical protein
	16079979	ytol	BSU29270	K	COG4109	hypothetical protein
	16080103	ytpA	BSU30510	I	COG2267	hypothetical protein
	16080100	ytqA	BSU30480	R	COG1242	hypothetical protein
	16080086	ytvA	BSU30340	T	COG2202	hypothetical protein
	16080055	ytzF	BSU30030	J	COG1187	hypothetical protein
	16080367	yvqJ	BSU33140	G	COG0477	hypothetical protein
	16080815	ywfM	BSU37630	G	COG0697	hypothetical protein
	16080757	ywkB	BSU37040	R	COG0679	hypothetical protein
	16080709	ywnH	BSU36560	M	COG1247	hypothetical protein
	16080681	ywqA	BSU36280	K	COG0553	hypothetical protein
	16080731	ywzB	BSU36780	S	COG4836	hypothetical protein
	16081109	yybO	BSU40570	G	COG0477	hypothetical protein
Sporulation Module B	16077125	yabM	BSU00570	R	COG2244	hypothetical protein
	16077127	yabO	BSU00590	J	COG1188	hypothetical protein
	16077129	yabQ	BSU00610	-	-	hypothetical protein
	16077133	yabS	BSU00650	R	COG2304	hypothetical protein
	16077225	ybaN	BSU01570	G	COG0726	hypothetical protein
	16077330	ycbR	BSU02610	-	-	hypothetical protein
	16077352	ycdF	BSU02830	Q	COG1028	hypothetical protein
	16077363	yceH	BSU02940	P	COG3853	hypothetical protein
	16077454	ycnD	BSU03860	C	COG0778	hypothetical protein
	16077533	ydcE	BSU04660	T	COG2337	hypothetical protein
	16077923	yfhK	BSU08570	-	-	hypothetical protein
	16077872	yfjL	BSU08050	-	-	hypothetical protein
	16077861	yfkC	BSU07940	M	COG0668	hypothetical protein
	16077855	yfkJ	BSU07880	T	COG0394	hypothetical protein
	16077836	yfiG	BSU07690	J	COG0024	hypothetical protein
	16078049	yhaU	BSU09850	P	COG0475	hypothetical protein
	16078044	yheB	BSU09790	S	COG4399	hypothetical protein
	16078041	yheE	BSU09760	-	-	hypothetical protein
	16078193	yjaU	BSU11280	-	-	hypothetical protein
	16078219	yjbG	BSU11540	E	COG1164	hypothetical protein
	16078221	yjbl	BSU11560	R	COG2346	hypothetical protein
	16078247	yjcD	BSU11820	L	COG0210	hypothetical protein
	16078192	yjzD	BSU11270	-	-	hypothetical protein
	16078373	ykkB	BSU13080	J	COG1670	hypothetical protein
	16078412	ykrK	BSU13480	-	-	hypothetical protein
	16078425	ykrY	BSU13610	G	COG0235	hypothetical protein
	16078531	yktC	BSU14670	G	COG0483	hypothetical protein
	16078485	ykuT	BSU14210	M	COG0668	hypothetical protein
	16078543	ylal	BSU14790	S	COG4896	hypothetical protein
	16078544	ylaJ	BSU14800	-	-	hypothetical protein
	16078559	ylbB	BSU14950	R	COG0517	hypothetical protein
	16078562	ylbE	BSU14980	-	-	hypothetical protein
	16078563	ylbF	BSU14990	S	COG3679	hypothetical protein
	16078574	ylbP	BSU15100	K	COG0454	hypothetical protein
	16078601	ylmD	BSU15370	S	COG1496	hypothetical protein
	16078752	ymfK	BSU16890	-	-	hypothetical protein
	16078839	yndE	BSU17760	-	-	hypothetical protein
	16078841	yndG	BSU17780	-	-	hypothetical protein
	16078847	yndN	BSU17840	E	COG0346	hypothetical protein
	16078854	yneF	BSU17910	S	COG3763	hypothetical protein
	16078859	yneK	BSU17960	-	-	hypothetical protein
	50812241	yneP	BSU18040	-	-	hypothetical protein
	16078867	yneR	BSU18060	S	COG4841	hypothetical protein
	16079176	yomZ	BSU21170	-	-	hypothetical protein
	16079054	yosZ	BSU19960	-	-	hypothetical protein
	16078934	yozG	BSU18740	K	COG3655	hypothetical protein
	16079351	ypdC	BSU22940	S	COG2339	hypothetical protein
	16079349	ypeB	BSU22920	-	-	hypothetical protein
	16079344	ypgA	BSU22870	C	COG1304	hypothetical protein

	Genbank ID	Gene Name	Gene ID	COG Fn.	COG No.	Description
	16079343	yphA	BSU22860	-	-	hypothetical protein
	16079339	yphE	BSU22820	-	-	hypothetical protein
	16079338	yphF	BSU22810	-	-	hypothetical protein
	16079316	yphA	BSU22590	R	COG0457	hypothetical protein
	16079286	yppD	BSU22280	-	-	hypothetical protein
	16079284	yppF	BSU22260	-	-	hypothetical protein
	16079279	yprB	BSU22210	L	COG3359	hypothetical protein
	16079680	yqaL	BSU26270	-	-	hypothetical protein
	16079591	yqfB	BSU25370	-	-	hypothetical protein
	16079590	yqfC	BSU25360	-	-	hypothetical protein
	16079568	yqfS	BSU25130	L	COG0648	hypothetical protein
	16079564	yqfW	BSU25090	-	-	hypothetical protein
	16079560	yqgA	BSU25050	-	-	hypothetical protein
	16079535	yqgX	BSU24790	R	COG0491	hypothetical protein
	16079507	yqhO	BSU24510	R	COG1752	hypothetical protein
	16079505	yqhQ	BSU24490	R	COG3872	hypothetical protein
	16079468	yqiQ	BSU24120	G	COG2513	hypothetical protein
	16079424	yqkA	BSU23670	S	COG2320	hypothetical protein
	16079823	yrbF	BSU27700	U	COG1862	hypothetical protein
	50812276	yrvl	BSU27590	-	-	hypothetical protein
	16079919	ysfB	BSU28670	K	COG3835	hypothetical protein
	16080010	ytbJ	BSU29580	H	COG0301	hypothetical protein
	16080137	ytdA	BSU30850	M	COG1210	hypothetical protein
	16080004	yteJ	BSU29520	S	COG1714	hypothetical protein
	16080003	ytfI	BSU29510	-	-	hypothetical protein
	16080116	ytkC	BSU30640	R	COG4824	hypothetical protein
	16080110	ytmA	BSU30580	E	COG1506	hypothetical protein
	16080044	ytmP	BSU29920	M	COG0510	hypothetical protein
	16079979	ytol	BSU29270	K	COG4109	hypothetical protein
	16080036	ytpP	BSU29840	O	COG0526	hypothetical protein
	16080097	ytrB	BSU30450	V	COG1131	hypothetical protein
	16080096	ytrC	BSU30440	-	-	hypothetical protein
	16079949	ytxB	BSU28970	S	COG0398	hypothetical protein
	16080566	yvlA	BSU35130	-	-	hypothetical protein
	16080367	yvqJ	BSU33140	G	COG0477	hypothetical protein
	16080576	yvzD	BSU35230	-	-	hypothetical protein
	16080847	ywdH	BSU37960	C	COG1012	hypothetical protein
	16080844	ywdK	BSU37930	S	COG2363	hypothetical protein
	16080751	ywlA	BSU36980	-	-	hypothetical protein
	16080636	ywtG	BSU35830	G	COG0477	hypothetical protein
	16080731	ywzB	BSU36780	S	COG4836	hypothetical protein
	16081039	yxbC	BSU39880	S	COG2850	hypothetical protein
	16080919	yxID	BSU38680	-	-	hypothetical protein
Sporulation Module C	16077125	yabM	BSU00570	R	COG2244	hypothetical protein
	16077127	yabO	BSU00590	J	COG1188	hypothetical protein
	16077129	yabQ	BSU00610	-	-	hypothetical protein
	16077133	yabS	BSU00650	R	COG2304	hypothetical protein
	16077330	ycbR	BSU02610	-	-	hypothetical protein
	16077533	ydcE	BSU04660	T	COG2337	hypothetical protein
	16078049	yhaU	BSU09850	P	COG0475	hypothetical protein
	16078041	yheE	BSU09760	-	-	hypothetical protein
	16078412	ykrK	BSU13480	-	-	hypothetical protein
	16078601	yImD	BSU15370	S	COG1496	hypothetical protein
	16078839	yndE	BSU17760	-	-	hypothetical protein
	16078841	yndG	BSU17780	-	-	hypothetical protein
	50812241	yneP	BSU18040	-	-	hypothetical protein
	16078933	yoaS	BSU18730	-	-	hypothetical protein
	16079176	yomZ	BSU21170	-	-	hypothetical protein
	16079054	yosZ	BSU19960	-	-	hypothetical protein
	16079351	ypdC	BSU22940	S	COG2339	hypothetical protein
	16079338	yphF	BSU22810	-	-	hypothetical protein
	16079591	yqfB	BSU25370	-	-	hypothetical protein
	16079590	yqfC	BSU25360	-	-	hypothetical protein
	16079823	yrbF	BSU27700	U	COG1862	hypothetical protein

Genbank ID	Gene Name	Gene ID	COG Fn.	COG No.	Description
16080010	ytbJ	BSU29580	H	COG0301	hypothetical protein
16080137	ytdA	BSU30850	M	COG1210	hypothetical protein
16080003	ytfI	BSU29510	-	-	hypothetical protein
16079949	ytxB	BSU28970	S	COG0398	hypothetical protein
16080576	yvzD	BSU35230	-	-	hypothetical protein
16080751	ywlA	BSU36980	-	-	hypothetical protein
16080636	ywtG	BSU35830	G	COG0477	hypothetical protein
16081039	yxbC	BSU39880	S	COG2850	hypothetical protein