Journal of
Molecular Microbiology
and Biotechnology

J Mol Microbiol Biotechnol 2007;12:165–179
DOI: 10.1159/000099639

Large-Scale Comparative Genomic Analyses of Cytoplasmic Membrane Transport Systems in Prokaryotes

Qinghu Ren Ian T. Paulsen

The Institute for Genomic Research, Rockville, Md., USA

Key Words
Membrane transporter · Prokaryotic genomes · Eubacteria · Archaea · Phylogenetic profiling

Abstract
The recent advancements in genome sequencing make it possible for the comparative analyses of essential cellular processes like transport in organisms across the three domains of life. Membrane transporters play crucial roles in fundamental cellular processes and functions in prokaryotic systems. Between 3 and 16% of open reading frames in prokaryotic genomes were predicted to encode membrane transport proteins, emphasizing the importance of transporters in their lifestyles. Hierarchical clustering of phylogenetic profiles of transporter families, which are derived from the presence or absence of a certain transporter family, showed distinct clustering patterns for obligate intracellular organisms, plant/soil-associated microbes and autotrophs. Obligate intracellular organisms possess the fewest types and number of transporters presumably due to their relatively stable living environment, while plant/soil-associated organisms generally encode the largest variety and number of transporters. A group of autotrophs are clustered together largely due to their absence of transporters for carbohydrate and organic nutrients and the presence of transporters for inorganic nutrients. Inside of each group, organisms are further clustered by their phylogenetic properties. These findings strongly suggest the correlation of transporter profiles to both evolutionary history and the overall physiology and lifestyles of the organisms.

Copyright © 2007 S. Karger AG, Basel

Introduction

Membrane transport systems are vital to every living organism. Transporters function in the acquisition of organic nutrients, maintenance of ion homeostasis, extrusion of toxic and waste compounds, environmental sensing and cell communication, and other important cellular functions [Saier, 1999], therefore playing essential roles in life-endowing processes like metabolism, communication, and reproduction. There has also been increasing evidence suggesting the relevance of the composition of membrane transport systems to the general physiology and lifestyles of the organisms [Paulsen et al., 1998, 2000; Ren and Paulsen, 2005].

Various transport systems differ in their putative membrane topology, energy coupling mechanisms and substrate specificities [Saier, 2000]. The most commonly utilized energy sources to drive transport are adenosine triphosphate (ATP), phosphoenolpyruvate, or chemiosmotic energy in the form of sodium ion or proton electrochemical gradients. Primary active transporters couple the transport process to a primary source of energy

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2007 S. Karger AG, Basel
1464–1801/07/0124–0165$23.50/0

Accessible online at:
www.karger.com/mmb

Ian T. Paulsen
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850 (USA)
Tel. +1 301 795 7531, Fax +1 301 838 0208, E-Mail ipaulsen@tigr.org

(ATP hydrolysis), for example, the MalKGFE maltose transporter from *Escherichia coli* [Bohm et al., 2002; Boos and Shuman, 1998]. Secondary transporters utilize an ion or solute electrochemical gradient, such as the proton/sodium motive force, to drive the transport process, e.g. *E. coli* LacY lactose permease [Abramson et al., 2003; Newman et al., 1981; Viitanen et al., 1986]. Group translocators transport and phosphorylate their substrates. *E. coli* MtlA mannitol PTS transporter phosphorylates exogenous mannitol using phosphoenolpyruvate as the phosphoryl donor and energy source and releases the phosphate ester, mannitol-1-P, into the cell cytoplasm [Elferink et al., 1990; Postma et al., 1993]. Compared to other transporter types, channels are unique in that they are energy-independent transporters that transport water, specific types of ions or hydrophilic small molecules down a concentration or electrical gradient with higher rates of transport and lower stereospecificity, e.g. *E. coli* GlpF glycerol channel [Sweet et al., 1990].

Cytoplasmic membrane transporters typically consist of at least one membrane-localized protein component with multiple transmembrane-spanning α-helical segments. This has led to membrane transport systems being difficult to study experimentally. The genomic/bioinformatic analyses provide an attractive alternative to study membrane transporters [Ren et al., 2004]. As of today, over 400 prokaryotic genomes have been sequenced and deposited in the public databases (Gold Genomes Online Database, http://www.genomesonline.org/) [Bernal et al., 2001; Janssen et al., 2005]. These genomes cover a broad range of microbial organisms from different phylogenetic groupings, allowing comparative genomic analyses across a diverse range of organisms and lifestyles. The functional prediction and classification of complete membrane transport systems in these sequenced genomes, as well as the comparative analyses of transporter profiles from related organisms are of great value in understanding organisms' physiology and lifestyles.

In this review, we present a comparative genomic study of prokaryotic membrane transport systems from 201 sequenced genomes, with the focus on their relationship to their overall physiology and lifestyles.

Comparative Genomic Analysis of Membrane Transport Systems

Bioinformatic analyses of 201 species, including 178 eubacteria and 23 archaea (table 1) enabled us to identify a total of 53,669 transport proteins. Based on sequence similarities and phylogenetic analyses, these transport proteins could be categorized into 94 families, including 5 families of primary transporters, 70 families of secondary transporters, 11 channel protein families, 2 phosphotransferase systems, and 6 unclassified families. Some of these families are very large superfamilies with numerous members, such as the ATP-binding cassette superfamily (ABC) and the major facilitator superfamily (MFS), both of which are widely distributed across the eubacterial and archaeal species. Some families, on the contrary, only exist in a very limited phylogenetic spectrum and/or are present in only limited numbers.

The total number of predicted cytoplasmic membrane transport proteins (fig. 1a) and the percentage of transport proteins relative to the total number of open reading frames (ORFs) (fig. 1b) were compared for the 201 prokaryotes (listed by their phylogenetic groupings). Between 3 and 16% of ORFs in prokaryotic genomes were predicted to encode membrane transport proteins, emphasizing the importance of transporters in the lifestyles of all eubacterial and archaeal species. There is considerable variation on the quantity of transport proteins, even for species within the same phylogenetic group. For example, organisms within the α-Proteobacteria exhibit distinct lifestyles and corresponding differences in transporter contents. They include the rhizosphere-dwelling organisms *Mesorhizobium loti* (884 transport proteins, 12.2% of ORFs), *Bradyrhizobium japonicum* (987, 11.9%) and *Sinorhizobium meliloti* (827, 13.3%); the plant pathogen *Agrobacterium tumefaciens* (824, 15.3%); the human pathogens *Brucella* spp. (360–379, 11.0–11.9%); marine Roseobacters, like *Silicibacter pomeroyi* (571, 13.4%) and *Jannaschia* sp. (507, 12.0%), and obligate intracellular pathogens or endosymbionts such as *Rickettsia* spp., *Wolbachia* spp., *Anaplasma* spp., and *Ehrlichia* spp. (53–59, 4.5–7.0%). Across all phyla, obligate endosymbionts and intracellular pathogens generally seem to possess the most limited repertoire of membrane transporters.

Organisms with the lowest percent of ORFs encoding transport proteins include *Pirellula* sp. (225, 3.1%), a marine aerobic heterotrophic planctomycete; *Leptospira interrogans* (147, 3.1%), a parasitic pathogenic spirochaete, and several archaeal species, such as *Methanococcus jannaschii* (68, 3.9%), *Methanopyrus kandleri* (54, 3.2%), and *Nanoarchaeum equitans* (17, 3.0%). One of the contributing factors could be the very limited experimental characterization of species in these phylogenetic groupings, which serves the base for bioinformatic predictions. Pre-

Table 1. Organisms used in this study and their transport proteins

Taxonomy	Organism name	Organism ID	Total transport proteins	Percent of ORFs (%)
Archaea-Crenarchaeota	*Aeropyrum pernix* K1	1	158	8.6
	Pyrobaculum aerophilum IM2	2	146	5.6
	Sulfolobus solfataricus P2	3	191	6.4
	Sulfolobus tokodaii strain7	4	166	5.9
	Sulfolobus acidocaldarius DSM639	5	152	6.8
Archaea-Euryarchaeota	*Archaeoglobus fulgidus* DSM4304	6	184	7.6
	Halobacterium sp. NRC-1	7	160	6.1
	Methanosarcina acetivorans C2A	8	394	8.7
	Methanococcus jannaschii DSM	9	68	3.9
	Methanopyrus kandleri AV19	10	54	3.2
	Methanococcus maripaludis S2	11	138	8.0
	Methanosarcina mazei Goe1	12	248	7.4
	Methanobacterium thermoautotrophicum ΔH	13	102	5.4
	Pyrococcus abyssi GE5	14	177	10.0
	Pyrococcus furiosus DSM3638	15	195	9.4
	Pyrococcus horikoshii OT3	16	159	8.8
	Picrophilus torridus DSM9790	17	171	11.1
	Thermoplasma acidophilum DSM1728	18	145	9.8
	Thermoplasma volcanium GSS1	19	143	4.7
	Haloarcula marismortui ATCC43049	20	330	7.8
	Natronomonas pharaonis DSM2160	21	216	7.7
	Thermococcus kodakaraensis KOD1	22	198	8.6
Archaea-Nanoarchaea	*Nanoarchaeum equitans* Kin4-M	23	17	3.0
Actinobacteria	*Bifidobacterium longum* NCC2705	24	233	13.5
	Corynebacterium diphtheriae NCTC13129	25	251	11.0
	Corynebacterium efficiens YS-314	26	303	10.3
	Corynebacterium glutamicum ATCC13032	27	355	11.9
	Leifsonia xyli CTCB07	28	179	8.8
	Mycobacterium avium K-10	29	293	6.7
	Mycobacterium bovis AF2122/97	30	240	6.1
	Mycobacterium leprae TN	31	100	6.2
	Mycobacterium tuberculosis H37Rv	32	238	6.1
	Nocardia farcinica IFM10152	33	443	7.5
	Propionibacterium acnes KPA171202	34	297	12.9
	Streptomyces avermitilis MA-4680	35	705	9.3
	Streptomyces coelicolor A3(2)	36	702	8.9
	Tropheryma whippelii TW08/27	37	64	8.2
	Tropheryma whipplei Twist	38	25	8.7
Aquificae	*Aquifex aeolicus* VF5	39	89	5.8
Bacteroidetes	*Bacteroides fragilis* YCH46	40	256	5.5
	Bacteroides fragilis NCTC9343	41	256	6.0
	Bacteroides thetaiotaomicron VPI-5482	42	253	5.3
Chlamydia	*Chlamydophila caviae* GPIC	44	76	7.6
	Chlamydia muridarum Nigg	45	66	7.2
	Chlamydia pneumoniae AR39	46	74	6.7
	Chlamydophila pneumoniae TW-183	47	73	6.6
	Chlamydia trachomatis serovar D	48	69	7.7
	Parachlamydia sp. UWE25	49	121	6.0
Chlorobi	*Chlorobium chlorochromatii* CaD3	50	114	5.7
	Chlorobium tepidum TLS	43	122	5.4
	Pelodictyon luteolum DSM273	51	164	7.9

Table 1 (continued)

Taxonomy	Organism name	Organism ID	Total transport proteins	Percent of ORFs (%)
Chloroflexi	*Dehalococcoides ethenogenes* 195	52	101	6.4
	Dehalococcoides sp. CBDB1	53	102	7.0
Cyanobacteria	*Gloeobacter violaceus* PCC7421	54	246	5.6
	Nostoc sp. PCC7120	55	382	6.2
	Prochlorococcus marinus MIT9313	56	142	6.3
	Prochlorococcus marinus SS120(CCMP1375)	57	88	4.7
	Prochlorococcus marinus MED4(CCMP1378)	58	94	5.5
	Synechococcus elongatus PCC6301	59	185	7.3
	Synechocystis sp. PCC6803	60	220	6.9
	Synechococcus sp. WH8102	61	148	5.9
	Thermosynechococcus elongatus BP-1	62	166	6.7
Deinococcus-Thermus	*Deinococcus radiodurans* R1	63	262	8.2
	Thermus thermophilus HB27	64	212	9.6
Firmicutes	*Bacillus anthracis* Ames	65	564	10.6
	Bacillus anthracis A2012	66	682	12.3
	Bacillus cereus ATCC14579	67	571	10.9
	Bacillus halodurans C-125	68	510	12.5
	Bacillus licheniformis ATCC14580	69	503	12.1
	Bacillus subtilis 168	70	423	10.3
	Bacillus thuringiensis konkukian 97-27	71	626	12.2
	Clostridium acetobutylicum ATCC824	72	371	9.6
	Carboxydothermus hydrogenoformans Z-2901	73	166	6.3
	Clostridium perfringens 13	74	311	11.4
	Clostridium tetani E88	75	266	11.2
	Enterococcus faecalis V583	76	393	12.6
	Geobacillus kaustophilus HTA426	77	319	9.0
	Lactobacillus acidophilus NCFM	78	268	14.4
	Listeria innocua Clip11262 (rhamnose-negative)	79	380	12.5
	Lactobacillus johnsonii NCC533	80	286	15.7
	Lactococcus lactis IL1403	81	245	10.8
	Listeria monocytogenes EGD-e	82	387	13.6
	Listeria monocytogenes 4b	83	370	13.1
	Lactobacillus plantarum WCFS1	84	401	13.3
	Mesoplasma florum L1	85	74	10.8
	Mycoplasma gallisepticum R	86	76	10.5
	Mycoplasma genitalium G-37	87	55	11.4
	Mycoplasma hyopneumoniae 232	88	94	13.6
	Mycoplasma mobile 163K	89	69	10.9
	Mycoplasma mycoides PG1T	90	103	10.1
	Mycoplasma penetrans HF-2	91	94	9.1
	Mycoplasma pneumoniae M129	92	47	6.9
	Mycoplasma pulmonis UAB CTIP	93	89	11.4
	Oceanobacillus iheyensis HTE831	94	439	12.6
	Phytoplasma asteris OY-M	95	55	7.3
	Streptococcus agalactiae 2603V/R	96	274	12.9
	Streptococcus agalactiae NEM316	97	278	13.3
	Staphylococcus aureus N315	98	324	12.3
	Staphylococcus aureus COL	99	276	10.5
	Staphylococcus epidermidis ATCC12228	100	274	11.3
	Staphylococcus epidermidis RP62a	101	269	10.6
	Streptococcus mutans UAB159	102	240	12.2
	Streptococcus pneumoniae TIGR4	103	261	12.5

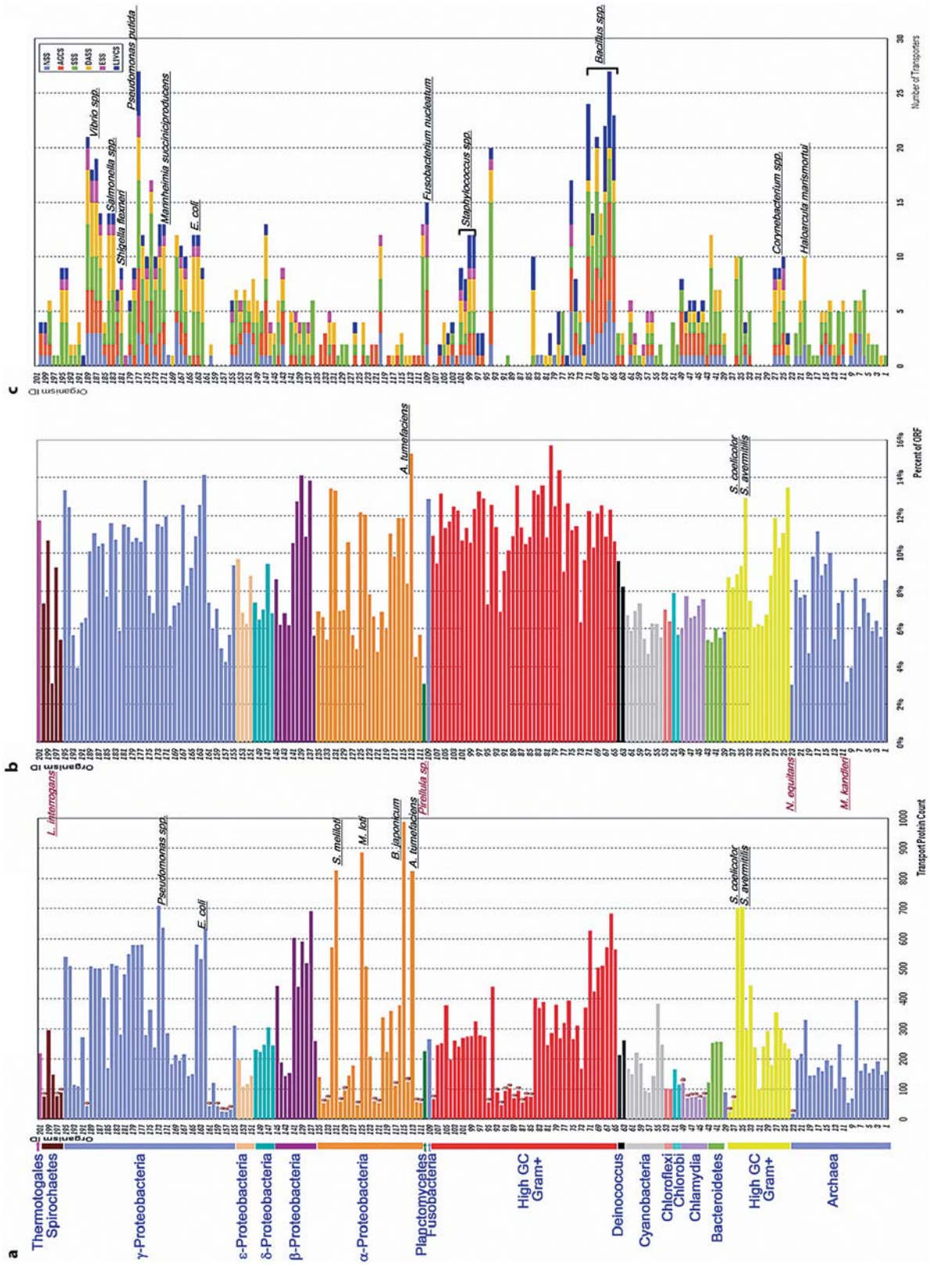
Table 1 (continued)

Taxonomy	Organism name	Organism ID	Total transport proteins	Percent of ORFs (%)
	Streptococcus pyogenes M1	104	198	11.7
	Symbiobacterium thermophilum IAM14863	105	378	11.3
	Streptococcus thermophilus CNRZ1066	106	252	13.2
	Thermoanaerobacter tengcongensis MB4	107	245	9.5
	Ureaplasma urealyticum serovar 3	108	67	10.9
Fusobacteria	*Fusobacterium nucleatum* ATCC25586	109	266	12.9
Planctomycetes	*Pirellula* sp. 1	110	225	3.1
α-Proteobacteria	*Anaplasma marginale* St. Maries	111	54	5.7
	Anaplasma phagocytophilum HZ	112	57	4.5
	Agrobacterium tumefaciens C58	113	824	15.3
	Bartonella henselae Houston-1	114	125	8.4
	Bradyrhizobium japonicum USDA110	115	987	11.9
	Brucella melitensis 16M	116	379	11.9
	Bartonella quintana Toulose	117	112	9.8
	Brucella suis 1330	118	360	11.0
	Caulobacter crescentus CB15	119	223	6.0
	Colwellia psychroerythraea 34H	120	338	6.9
	Ehrlichia chaffeensis Arkansas	121	53	4.8
	Ehrlichia ruminantium Welgevonden	122	59	6.6
	Gluconobacter oxydans 621H	123	208	7.8
	Jannaschia sp. CCS1	124	507	12.0
	Mesorhizobium loti MAFF303099	125	884	12.2
	Neorickettsia sennetsu Miyayama	126	46	4.9
	Nitrobacter winogradskyi Nb-255	127	177	5.7
	Candidatus Pelagibacter ubique HTCC1062	128	143	10.6
	Rickettsia conorii Malish7	129	96	7.0
	Rickettsia prowazekii MadridE	130	58	6.9
	Sinorhizobium meliloti 1021	131	827	13.3
	Silicibacter pomeroyi DSS-3	132	571	13.4
	Wolbachia pipientis wMel	133	65	5.4
	Wolbachia sp. TRS *(Brugia malayi)*	134	53	6.6
	Zymomonas mobilis ZM4	135	138	6.9
β-Proteobacteria	*Azoarcus* sp. EbN1	136	259	5.6
	Bordetella bronchiseptica RB50 NCTC-13252	137	691	13.8
	Burkholderia mallei ATCC23344	138	518	10.9
	Bordetella parapertussis 12822 NCTC-13253	139	590	14.1
	Bordetella pertussis Tohama I NCTC-13251	140	439	12.7
	Burkholderia pseudomallei K96243	141	603	10.5
	Nitrosomonas europaea ATCC19718	142	152	6.2
	Neisseria meningitidis MC58	143	142	6.8
	Nitrosococcus oceani ATCC19707	144	187	6.2
	Ralstonia solanacearum GMI1000	145	441	8.6
δ-Proteobacteria	*Bdellovibrio bacteriovorus* HD100	146	244	6.8
	Desulfotalea psychrophila LSv54	147	305	9.4
	Desulfovibrio vulgaris Hildenborough	148	247	7.0
	Geobacter sulfurreducens PCA	149	223	6.5
	Pelobacter carbinolicus DSM2380	150	230	7.4
ε-Proteobacteria	*Campylobacter jejuni* NCTC11168	151	144	8.8
	Helicobacter hepaticus ATCC51449	152	117	6.2
	Helicobacter pylori 26695	153	108	6.9
	Wolinella succinogenes	154	198	9.7

Table 1 (continued)

Taxonomy	Organism name	Organism ID	Total transport proteins	Percent of ORFs (%)
γ-Proteobacteria	*Acinetobacter* sp. ADP1	155	311	9.4
	Buchnera aphidicola Sg	156	31	5.7
	Buchnera aphidicola APS	157	24	4.3
	Buchnera aphidicola Bp	158	25	5.0
	Candidatus Blochmannia pennsylvanicus BPEN	159	43	7.0
	Coxiella burnetii RSA493	160	121	6.0
	Candidatus Blochmannia floridanus	161	43	7.4
	Erwinia carotovora SCRI1043	162	632	14.1
	Escherichia coli K12-MG1655	163	532	12.6
	Escherichia coli O157:H7 EDL933	164	580	10.9
	Francisella tularensis Schu4	165	148	9.2
	Haemophilus ducreyi 35000HP	166	142	8.3
	Haemophilus influenzae KW20	167	215	12.5
	Idiomarina loihiensis L2TR	168	194	7.4
	Legionella pneumophila Philadelphia 1	169	212	7.2
	Methylococcus capsulatus Bath	170	182	6.2
	Mannheimia succiniciproducens MBEL55E	171	285	12.0
	Pseudomonas aeruginosa PAO1	172	635	11.4
	Pseudomonas fluorescens Pf-5	173	708	11.5
	Pseudoalteromonas haloplanktis TAC125	174	238	6.8
	Photorhabdus lumines laumondii TTO1	175	363	7.8
	Photobacterium profundum SS9	177	580	10.6
	Pseudomonas syringae pv. tomato DC3000	179	579	10.6
	Rhodopseudomonas palustris CGA009	180	548	11.4
	Shigella flexneri 2a 301	181	481	11.5
	Shewanella oneidensis MR-1	182	281	5.9
	Salmonella typhi CT18	183	510	10.7
	Salmonella typhimurium LT2	184	516	11.6
	Thiomicrospira crunogena XCL-2	185	169	7.7
	Vibrio cholerae El Tor N16961	186	403	10.5
	Vibrio parahaemolyticus RIMD2210633	187	500	10.3
	Vibrio vulnificus CMCP6	188	501	11.0
	Vibrio vulnificus YJ016	189	507	10.1
	Wigglesworthia glossinidia P-endosymbiont	190	43	6.6
	Xanthomonas axonopodis pv. citri 306	191	272	6.3
	Xylella fastidiosa 9a5c	192	109	3.9
	Xylella fastidiosa Temecula1	193	115	5.7
	Yersinia pestis CO-92	194	508	12.4
	Yersinia pseudotuberculosis IP32953	195	539	13.3
Spirochaetes	*Borrelia burgdorferi* B31	196	89	5.4
	Borrelia garinii PB1	197	77	9.3
	Leptospira interrogans serovar lai56601	198	147	3.1
	Treponema denticola ATCC35405	199	295	10.7
	Treponema pallidum Nichols	200	76	7.3
Thermotogales	*Thermotoga maritima* MSB8	201	218	11.7

Fig. 1. The overall numbers of recognized transport proteins. Organisms from distinct phylogenetic groups are labeled with different colors. The obligate intracellular parasites/pathogens are marked with red stars. **a** Total number of transport proteins in 201 prokaryotes. **b** Transport proteins as the percentage of total ORFs. **c** Distribution of sodium-dependent amino acid/solute symporters across six families: NSS = light blue; AGCS = orange; SSS = salmon; DASS = lime; glutamate:sodium symporter family (ESS) = pink; LIVCS = dark blue.


a
b
c
Thermotogales
Spirochaetes
γ-Proteobacteria
ε-Proteobacteria
δ-Proteobacteria
β-Proteobacteria
α-Proteobacteria
Planctomycetes
Fusobacteria
High GC Gram+
Deinococcus
Cyanobacteria
Chloroflexi
Chlorobi
Chlamydia
Bacteroidetes
High GC Gram+
Archaea
L. interrogans
Pseudomonas spp.
E. coli
S. meliloti
M. loti
B. japonicum
A. tumefaciens
Pirellula sp.
S. coelicolor
S. avermitilis
N. equitans
M. kandleri
Transport Protein Count
Percent of ORF
Organism ID
Vibrio spp.
Salmonella spp.
Shigella flexneri
Mannheimia succiniciproducens
Pseudomonas putida
Fusobacterium nucleatum
Staphylococcus spp.
Bacillus spp.
Corynebacterium spp.
Haloarcula marismortui
NSS
AGCS
SSS
DASS
ESS
BLIVCS
Number of Transporters

vious studies show that archaeal species have much higher percentages of membrane proteins assigned to the role category of 'hypothetical proteins' than eubacterial species [Ren and Paulsen, 2005]. Some of these 'hypothetical proteins' could function in novel transport processes.

Although obligate intracellular organisms and small free-living parasites overall present the fewest transport proteins, they still devote a relatively high percentage of ORFs towards transport functions. For example, *Mycoplasma* spp. have over 10% of their ORFs encoding transport proteins. Transport proteins consist of average 7.7% of ORFs in 38 obligate intracellular and small free-living parasites, compared to an average of 8.9% in all organisms. Although most of these organisms appear to have undergone substantial reductive evolution, it seems that they have not preferentially lost or retained transporter genes as a consequence of their adaptation to intracellular lifestyles. Most of these organisms have various defective biosynthesis pathways, and have to uptake essential nutrients and intermediate metabolites from their host. Detailed examinations of transporter profiles show that these organisms have different degrees of reduction as to the types of transporters and categories of substrate. Compared to other prokaryotes, obligate intracellular organisms exhibit greater degree of reduction in efflux pumps and transporters for ions and small inorganic compounds. However, they appear to have retained a significant percentage of importers in the genome for essential nutrients and intermediate metabolites.

Phylogenetic Profiling as a Tool for Investigating Membrane Transport Content

The phylogenetic profile of a gene is a pattern representing the presence or absence of homologues in a set of fully sequenced genomes. Genes with similar phylogenetic profiles, as assessed by Pearson correlation coefficient, likely could function together in a pathway or are part of a complex because they are likely to evolve in a correlated fashion and tend to be either preserved or eliminated during evolution [Pellegrini et al., 1999; Pellegrini, 2001]. Phylogenetic profiling has many applications in genomics studies, such as detection of conserved core genes, lineage-specific gene family expansions [Vandepoele and Van de Peer, 2005], subcellular localization of proteins [Marcotte et al., 2000], prediction of physical and functional interactions and deduction of the functions of genes that have no well-characterized homologues [Levesque et al., 2003; Wu et al., 2005].

Previously we employed a novel application of phylogenetic profiling to investigate the presence or absence of transporter protein families across 141 sequenced prokaryotes and eukaryotes [Ren and Paulsen, 2005]. Compared to other studies, we used protein families rather than individual proteins as the unit of comparison. The phylogenetic profiling of transporter families provided interesting insights into the distribution of transporters across a broad range of organisms. Organisms from various phylogenetic groups which are adapted to similar environmental niches were often found in clusters. Inside each cluster, organisms were further grouped together by their phylogenetic history. Given that the profiling approach solely utilizes presence/absence of a transporter family and does not use sequence similarity directly, these findings suggest that the types of transporters utilized by an organism are related both to their physiology and to their evolutionary history.

The fast growing number of completely sequenced genomes enabled us to enhance the resolution of this phylogenetic profiling analysis. With the data on membrane transport systems from 201 fully sequenced prokaryotes, we were able to construct more detailed phylogenetic profiles for each transporter family (fig. 2). In line with our previous observations, hierarchical clustering of phylogenetic profiles showed a strong correlation between the observed clustering pattern and phylogeny, with distinct phylogenetic groupings of eubacteria and archaea clearly separated into different clusters, such as high GC Gram+, low GC Gram+, Proteobacteria, Chlamydia, Cyanobacteria, etc. (fig. 2, 3). Additionally, the clustering patterns are influenced by the lifestyle of organisms. The obligate intracellular pathogens/symbionts, the soil/plant-associated microbes and a collection of autotrophs are separated into distinct super-clusters (fig. 2, 3). These findings demonstrate that phylogenetic profiling is a viable and potent approach to the bioinformatic study of membrane transporters.

The obligate intracellular pathogens/symbionts cluster includes a group of phylogenetically diverse organisms (fig. 3b), including Chlamydia (pathogens); γ-Proteobacteria like *Buchnera* spp., *Wigglesworthia glossinidia* and *Candidatus Blochmannia* spp. (endosymbionts); α-Proteobacteria such as *Wolbachia* spp. (endosymbionts) and *Anaplasma* spp., *Ehrlichia* spp., *Rickettsia* spp., *Neorickettsia sennetsu*, *Bartonella* spp. (pathogens); low GC Gram+-like organisms *Mycoplasma* spp., *Ureaplasma urealyticum, Phytoplasma asteris* and *Tropheryma whipplei* (pathogens); Spirochetes like *Treponema pallidum*, *Borrelia* spp. (pathogens); and an archaeal endo-

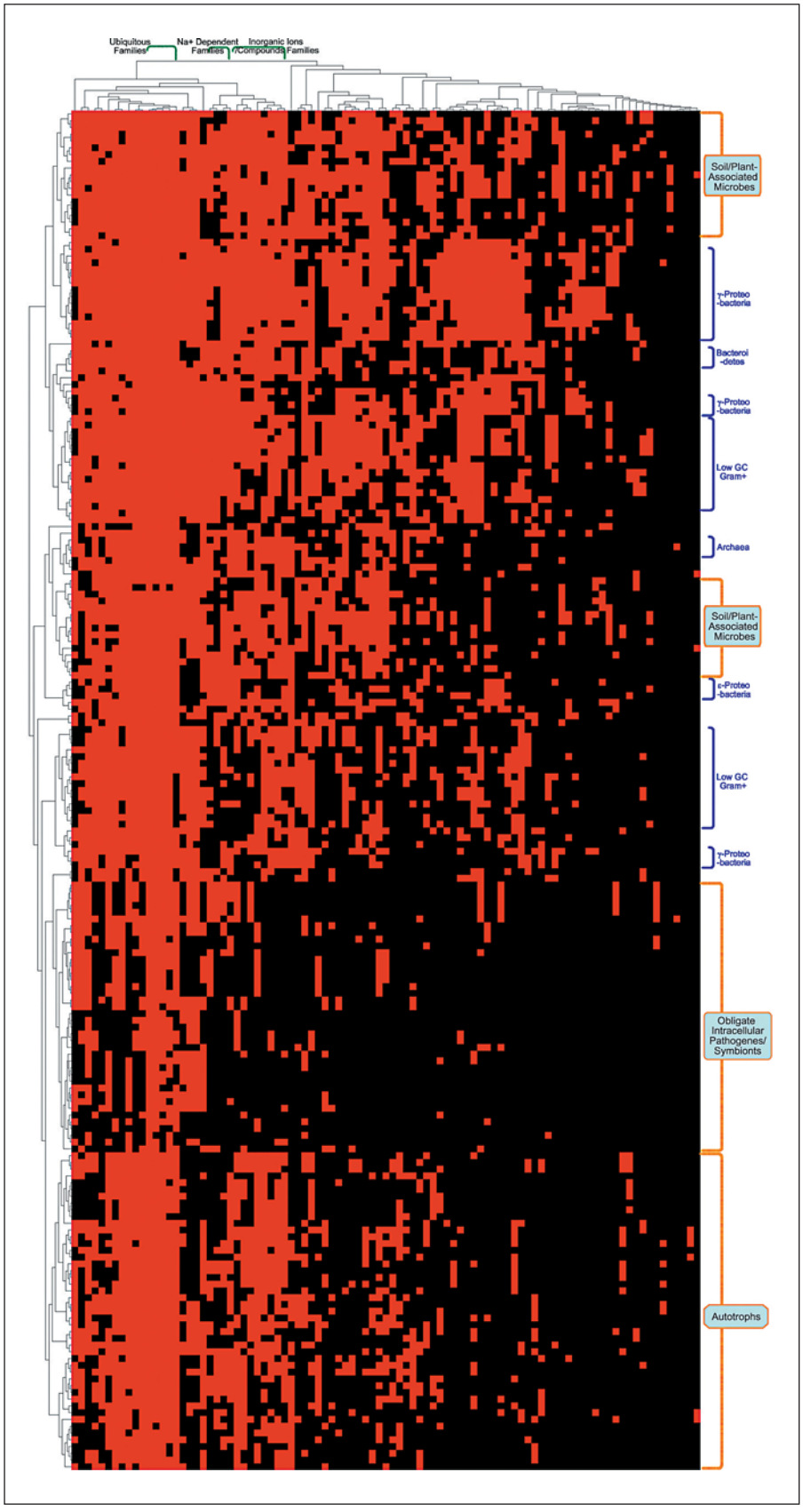


Fig. 2. Phylogenetic profiling of transporter families. Phylogenetic profiles were created for each transporter family. Each profile is a string with 201 entries (number of organisms analyzed). If a given family is present in an organism, the value '1' is assigned at this position (red color). If not, '0' is assigned (black color). Organisms and transporter families were clustered according to the similarity of their phylogenetic profiles.

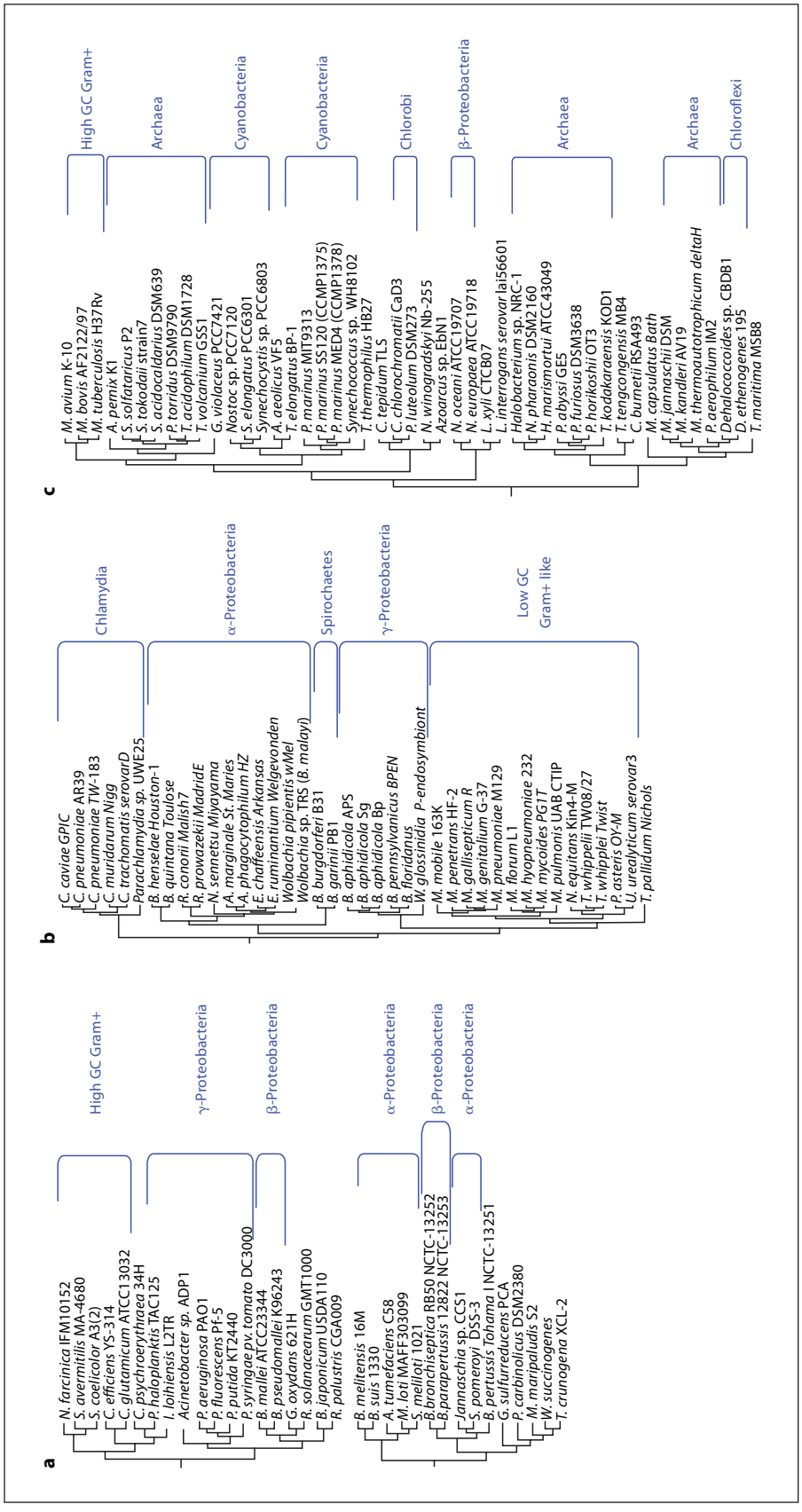


Fig. 3. Detailed view of three clusters of organisms generated by hierarchical clustering of their phylogenetic profiles of transporter families: soil/plant-associated microbes (**a**), obligate intracellular pathogens/symbionts (**b**) and autotrophs (**c**).

symbiont, *N. equitans*. Organisms in this cluster are mostly obligate intracellular organisms, with one or two exceptions, e.g. *Bartonella* spp. that are facultative intracellular pathogens. The transport needs for these obligate intracellular organisms are probably more specialized than those of environmental organisms due to the less dynamic nature of their intracellular environments. This may have allowed them to shed, for example, transporters for alternative nitrogen/carbon sources, drug/toxic metabolite efflux, osmoregulation, and ion homeostasis. The residual transport systems conserved in these obligate intracellular organisms probably belong to the core essential genes required for the acquisition of key nutrients and metabolic intermediates. For example, in *Rickettsia* species, genes coding for proteins functioning in glycolysis and the biosynthesis of S-adenosylmethionine and nucleotides are absent [Andersson and Andersson, 1999; Andersson et al., 1998; Dunning Hotopp et al., 2006; Ogata et al., 2001]. They completely rely on the hosts for these small molecules. As expected, transporter systems for the uptake of nucleoside monophosphates (ATP:ADP antiporter family), S-adenosylmethionine (drug/metabolite transporter superfamily) [Tucker et al., 2003], and glycerol-3-phosphate (MFS family) have been identified [Dunning Hotopp et al., 2006]. The essential glutamate transporters in two obligate endosymbionts *Candidatus Blochmannia floridanus* and *W. glossinidia* provides another example: The GltP glutamate:proton symporter (DAACS family) is encoded in *B. floridanus* [Tolner et al., 1995], while the GltJKL ABC transporter is expressed in *W. glossinidia* [Linton and Higgins, 1998]. Both of these organisms have a truncated TCA cycle which begins with α-ketoglutarate and ends with oxaloacetate [Zientz et al., 2004]. Their TCA cycle could be closed by the transamination of the imported glutamate to aspartate, catalyzed by an aspartate aminotransferase which uses oxaloacetate as a cosubstrate and produces α-ketoglutarate. Compared to the plant/soil-associated microbes, obligate intracellular organisms show a higher degree of variation in terms of energy coupling mechanism and transport mode. These variations may reflect the unique internal environment inside the host cells. All these observations illustrate how adaptation of an organism to certain living conditions leads to changes in its transporter repertoire and at the same time determine the set of transporters that the organism cannot afford to lose.

The soil/plant-associated microbes form two clusters, including organisms from various phylogenetic groups (fig. 3a). The first cluster includes Actinobacteria (*Corynebacterium* spp., *Nocardia farcinica* and *Streptomyces* spp.), γ-Proteobacteria *(Actinobacter* sp., *Idiomarina loihiensis*, *Pseudomonas* spp., *Pseudoalteromonas haloplanktis* and *Rhodopseudomonas palustris)*, and β-Proteobacteria *(Burkholderia* and *Ralstonia)*. The second one includes mainly α-Proteobacteria *(A. tumefaciens, Brucella* spp., *Jannaschia*sp., *M. loti, S. pomeroyi* and *S. meliloti)*, β-Proteobacteria (*Bordetella* spp.), δ-Proteobacteria *(Geobacter sulfurreducens* and *Pelobacter carbinolicus)*, and ε-Proteobacteria *(Wolinella succinogenes)*. This is in contrast to our previous analysis [Ren and Paulsen, 2005] in which these organisms formed one coherent cluster with two major branches comparable to the two clusters shown here. The first cluster is close to other γ-Proteobacteria like *E. coli* which has the highest diversity of transporter families among all prokaryotic organisms, partly due to the extensive experimental studies carried on this model organism. The second cluster is close to other δ-Proteobacteria and ε-Proteobacteria. Therefore, the respective phylogenetic relationships of these two clusters override the linkage by the influence of living environment on transporter contents as observed previously. One of the possible reasons causing the disparity could be the great expansion of γ-Proteobacteria species used in this study which may have exerted a stronger influence on the clustering. All of the organisms in these two clusters possess a robust collection of transporter systems. The similarity of phylogenetic profiles of organisms in these clusters probably reflects the versatility of these organisms and their exposure to a wide range of different substrates in their natural environment. The majority of species in this cluster can be free-living in the soil and some are capable of living in a diverse range of environments. They generally share a broad range of transport capabilities for plant-derived compounds specifically and for organic nutrients in general. Interestingly, some of the human facultative pathogens, such as *Bordetella* and *Brucella*, are also grouped in this cluster. These pathogens have close relatives that are soil or plant-associated environmental organisms [Parkhill et al., 2003; Paulsen et al., 2002], so their transport capabilities probably reflect a combination of their evolutionary heritage, original environmental niche and their current transport needs.

The third significant cluster of phylogenetic profiling of transporter families consists primarily of autotrophs (fig. 3c). This cluster was not found in our previous analysis [Ren and Paulsen, 2005] because of the lack of data on autotrophs. Obligate autotrophs obtain energy exclusively by the oxidation of inorganic substrates and use

CO_2 as the only resource of carbon [Kowalchuk and Stephen, 2001], such as the nitrifying bacteria *Nitrobacter winogradskyi* (oxidizing nitrite ion); *Nitrosomonas europaea* and *Nitrosococcus oceani* (oxidizing ammonium ion). Facultative autotrophs obtain some part of their energy from oxidation of iron, sulfur, hydrogen, nitrogen, and carbon monoxide. These include green sulfur bacteria *(Chlorobium* spp. and *Pelodictyon luteolum)*, green nonsulfur bacteria (*Dehalococcoides* spp.), both of which are anaerobic photosynthetic bacteria; Cyanobacteria *(Prochlorococcus* spp., *Synechococcus* spp., *Synechocystis* sp., *Nostoc* sp., *Gloeobacter violaceus* and *Thermosynechococcus elongates)* which are aerobic photosynthetic bacteria; a hydrogen-oxidizing, microaerophilic, obligate chemolithoautotrophs *(Aquifex aeolicus)*; an obligate methanotroph, *Methylococcus capsulatus*; and a group of autotrophic archaeal species (*Aeropyrum pernix, Sulfolobus* spp., *Picrophilus torridus, Thermoplasma* spp., *Methanobacterium thermoautotrophicum, M. kandleri, M. jannaschii, Pyrobaculum aerophilum, Pyrococcus* spp., *Thermococcus kodakaraensis, Natronomonas pharaonis, Haloarcula marismortui,* and *Halobacterium* sp.). In line with their metabolism features, organisms in this cluster generally lack transporters for carbohydrates, amino acids, carboxylates and nucleosides, etc. Instead, they encode a full array of transporters for various cations and anions, ammonium, inorganic phosphate, and sulfate which feed into their autotrophic metabolism. These features distinguish this group of autotrophs from organisms in the plant/soil-associated and intracellular pathogen/endosymbiont clusters. Interestingly, some heterotrophic bacteria were included in this cluster. They generally fall into several categories: Pathogens that are evolved from environmental organisms, like *Leifsonia xyli* and *L. interrogans*; organisms with extensive ion transport systems and/or few organic nutrient transporters, like *Thermoanaerobacter tengcongensis, Coxiella burnetii* and *Mycobacterium* spp., and a Thermotogales *(Thermotoga maritima)* with extensive array of archaeal-lineage genes [Nelson et al., 1999], and was found to cluster with the archaeal species in this super-cluster.

Comparison of the transporter profiles of marine microbes shows a close relationship between their transporter profiles and their physiology and ecological niches. The sequenced marine microbes to date can be categorized into three groups according to their metabolism and ecological niche: Cyanobacteria clade (photosynthetic autotrophs); Roseobacter clade (such as *Jannaschia* sp. and *S. pomeroyi*) that are metabolically versatile and capable of utilizing diverse organic and inorganic nutrients in the coastal and oceanic planktonic environment [Moran et al., 2004]; and a group of oligotrophic bacteria that are metabolically conservative and more specialized in scavenging organic nutrients in seawater [Button, 1991], such as *Oceanobacillus iheyensis, Vibrio vulnificus, I. loihiensis, Pelagibacter ubique* and *Photobacterium profundum*.

Cyanobacteria, which feature few importers for organic nutrients and a more substantial array of transporters for ion and inorganic compounds, belong to the autotroph cluster (fig. 3). Detailed examination of two Cyanobacteria species with different ocean environmental niches shows quite different transporter profiles [Palenik et al., 2006]: the coastal cyanobacterium, *Synechococcus* sp. strain CC9311, has a much larger capacity to transport, store, utilize or export metals, especially iron and copper than an open ocean oligotrophic strain, *Synechococcus* sp. strain WH8102, which could be related to its greater capacity to sense and respond to changes in its (coastal) environment. In contrast, WH8102 has systems predicted for the efflux of arsenite and chromate [Palenik et al., 2003] that are not found in CC9311. The Roseobacter clade, however, was clustered with plant/soil-associated clusters (fig. 3) due to their abundant and diverse transporters for both organic nutrients (peptides, amino acids, sugars, putrescine and spermidine, taurine, glycine betaine and dimethylsulphoniopropionate, etc.) and inorganic compounds (urea, phosphate, inorganic ions, sulfate, etc.) which enable them to take advantage of transient occurrences of high-nutrient niches within a bulk low-nutrient environment. One of the distinguishing features of Roseobacters are their uncommonly high number of TRAP transporter systems (26 systems for *S. pomeroyi* and 28 for *Jannaschia* sp., no other sequenced genome has more), probably reflecting their capability to import carboxylic acids produced in surface waters during photo-oxidation of dissolved organic matters, like glyoxylate and acetate [Moran et al., 2004]. The metabolically conservative marine heterotrophs did not form any distinct grouping and were clustered primarily by their phylogenetic traits. For example, *O. iheyensis* was clustered with other *Bacillus* spp.; and *V. vulnificus* and *P. profundum* were clustered with other *Vibrio* spp. *P. ubique* represents one of the smallest free-living nonparasitic microorganism [Giovannoni et al., 2005] with 1,354 ORFs, of which 143 encode transport proteins (10.6%). Compared to obligate intracellular organisms, it encodes a large number of transporters for diverse ni-

trogenous compounds, such as ammonium, urea, basic amino acids, spermidine, and putrescine. These features clearly exclude it from the obligate intracellular organism cluster.

The clustering of transporter families also show features related to the lifestyles of organisms. The ubiquitous families, like ABC, MFS, P(F)-type ATPase, which are present in virtually every organism we analyzed, are clustered together. A group of sodium ion-dependent transporter families, the neurotransmitter:sodium symporter (NSS), alanine/glycine:cation symporter (AGCS), solute:sodium symporter (SSS), and divalent anion:sodium symporter (DASS) are clustered together. Transporters in these families are all symporters which utilize the sodium ion gradient to transport amino acid, solute, and/or divalent ions to cytoplasm. This clustering may suggest that these families co-occur in a specific set of organisms, presumably those most reliant on sodium ion-driven transport. Figure 1d shows the detailed distribution of six sodium ion-dependent amino acid/solute transporter families in the 201 prokaryotic organisms we analyzed. We see considerable variation in the distribution of these families among phylogenetically related species. For example, *Mycobacterium* spp. and *Corynebacterium* spp. are closely related Actinobacteria. *M. tuberculosis* and *C. diphtheriae* are both pathogens of human respiratory systems. *Corynebacterium* spp. encode members of all six sodium-dependent transporter families, while *Mycobacterium* spp. have none. In fact, *Mycobacterium* spp. were clustered with the autotrophic bacteria as an artifact on our phylogenetic profiling studies (fig. 3c) at least in part due to their lack of sodium-dependent transporters.

In general, environmental organisms such as *Bacillus* spp. (including *Oceanbacillus* and *Lactobacillus* spp.), *Colwellia psychroerythraea*, *Pirellula* sp. and *Pseudomonas* spp. present the highest number of sodium-dependent pumps, while organisms with autotrophic lifestyles encode very few sodium ion-driven transporters, and those they do possess are more likely involved in the uptake of simple compounds such as sulfate rather than amino acids or carboxylates. Some of these autotrophs completely lack this type of transporters, such as *Dehalococcoides* spp., *M. kandleri, N. winogradskyi,* and *N. europaea*. There are a couple of interesting exceptions: *H. marismortui*, a halophilic microorganism that thrives in extreme saline environments, encodes 10 members of sodium-dependent transporters in NSS, SSS and DASS families, the highest number among all archaeal species studied. *Halobacterium* sp., another archaeal organism, which, like *H. marismortui*, proliferates in saturating salt solutions, also has 6 members of such transporters. These probably reflect their adaptation to a high-salt environment. A group of human pathogens encode relatively large numbers of sodium-dependent pumps, including Enterobacteriaceae (such as *E. coli*, *Salmonella*, *Shigella*, *Yersinia* and *Vibrio* spp.), *Staphylococcus* spp., *Corynebacterium* spp. and *Fusobacterium nucleatum*, etc. These could also be related to the high-salt environment in human GI tract, oral cavity and respiratory tract. Actually human epithelial cells utilize the same mechanism to uptake nutrients from the GI tract and to regulate the internal homeostasis. Among those organisms with obligate intracellular lifestyles, which need to obtain nutrients like amino acids from their host, the majority do not encode sodium-dependent amino acid transporters. Instead, they typically encode ABC family amino acid transporter and/or APC family amino acid:proton symporters or amino acid:amino acid antiporters. *Chlamydia* spp. are the only group of obligate intracellular organisms that show homologues in each of the six sodium-dependent amino acid/solute families.

There is another cluster of families for inorganic ions and small compounds, including potassium and chloride ion channels, ammonium transporter, inorganic phosphate transporter, sulfate permease, and calcium: cation antiporter. Autotrophic eubacterial and archaeal organisms generally utilize transporters in these families for the uptake of inorganic compounds, as well as soil/plant-associated microbes and other environmental organisms. As expected, obligate intracellular pathogens and endosymbionts generally lack this type of transporters.

Conclusion

The era of complete genome sequencing has opened new horizons in our understanding of complex biological questions. Comparative genomic approaches for the analysis of membrane transport systems have provided us invaluable insights on how microbes adapt to their environment. The observations that organisms with similar lifestyles and/or ecologic niches (obligate intracellular, soil/plant-associated, or autotrophic) display similar phylogenetic profiles despite their phylogenetic differences strongly suggest the influence of their environments on their membrane transport gene complement.

References

Abramson J, Smirnova I, Kasho V, Verner G, Iwata S, Kaback HR: The lactose permease of *Escherichia coli*: overall structure, the sugar-binding site and the alternating access model for transport. FEBS Lett 2003;555:96–101.

Andersson JO, Andersson SG: Genome degradation is an ongoing process in *Rickettsia*. Mol Biol Evol 1999;16:1178–1191.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 1998;396:133–140.

Bernal A, Ear U, Kyrpides N: Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucl Acids Res 2001;29:126–127.

Bohm A, Diez J, Diederichs K, Welte W, Boos W: Structural model of MalK, the ABC subunit of the maltose transporter of *Escherichia coli*: implications for *mal* gene regulation, inducer exclusion, and subunit assembly. J Biol Chem 2002;277:3708–3717.

Boos W, Shuman H: Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. Microbiol Mol Biol Rev 1998;62:204–229.

Button DK: Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the Michaelis constant. Appl Environ Microbiol 1991;57:2033–2038.

Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen J, Seshadri R, Ren Q, Wu M, Utterback TR, Smith S, Lewis M, Khouri H, Zhang C, Niu H, Lin Q, Ohashi N, Zhi N, Nelson W, Brinkac LM, Dodson RJ, Rosovitz MJ, Sundaram J, Daugherty SC, Davidsen T, Durkin AS, Gwinn M, Haft DH, Selengut JD, Sullivan SA, Zafar N, Zhou L, Benahmed F, Forberger H, Halpin R, Mulligan S, Robinson J, White O, Rikihisa Y, Tettelin H: Comparative genomics of emerging human *Ehrlichiosis* agents. PLoS Genet 2006;2:e21.

Elferink MG, Driessen AJ, Robillard GT: Functional reconstitution of the purified phosphoenolpyruvate-dependent mannitol-specific transport system of *Escherichia coli* in phospholipid vesicles: coupling between transport and phosphorylation. J Bacteriol 1990;172:7119–7125.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappe MS, Short JM, Carrington JC, Mathur EJ: Genome streamlining in a cosmopolitan oceanic bacterium. Science 2005;309:1242–1245.

Janssen P, Goldovsky L, Kunin V, Darzentas N, Ouzounis CA: Genome coverage, literally speaking: the challenge of annotating 200 genomes with 4 million publications. EMBO Reports 2005;6:397–399.

Kowalchuk GA, Stephen JR: Ammonia-oxidizing bacteria: a model for molecular microbial ecology. Annu Rev Microbiol 2001;55:485–529.

Levesque M, Shasha D, Kim W, Surette MG, Benfey PN: Trait-to-gene: a computational method for predicting the function of uncharacterized genes. Curr Biol 2003;13:129–133.

Linton KJ, Higgins CF: The *Escherichia coli* ATP-binding cassette (ABC) proteins. Mol Microbiol 1998;28:5–13.

Marcotte EM, Xenarios I, van der Bliek AM, Eisenberg D: Localizing proteins in the cell from their phylogenetic profiles. Proc Natl Acad Sci USA 2000;97:12115–12120.

Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen JA, Rahe E, Sheldon WM, Ye W, Miller TR, Carlton J, Rasko DA, Paulsen IT, Ren Q, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Rosovitz MJ, Haft DH, Selengut J, Ward N: Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature 2004;432:910–913.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 1999;399:323–329.

Newman MJ, Foster DL, Wilson TH, Kaback HR: Purification and reconstitution of functional lactose carrier from *Escherichia coli*. J Biol Chem 1981;256:11804–11808.

Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM, Raoult D: Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 2001;293:2093–2098.

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterbury J: The genome of a motile marine *Synechococcus*. Nature 2003;424:1037–1042.

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R, Paulsen IT: Genome sequence of *Synechococcus* CC9311:Insights into adaptation to a coastal environment. Proc Natl Acad Sci USA 2006;103:13555–13559.

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MTG, Churcher CM, Bentley SD, Mungall KL, Cerdeno-Tarraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, Whitehead S, Barrell BG, Maskell DJ: Comparative analysis of the genome sequences of *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet 2003;35:32–40.

Paulsen IT, Nguyen L, Sliwinski MK, Rabus R, Saier MH Jr: Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. J Mol Biol 2000;301:75–100.

Paulsen IT, Seshadri R, Nelson KE, Eisen JA, Heidelberg JF, Read TD, Dodson RJ, Umayam L, Brinkac LM, Beanan MJ, Daugherty SC, Deboy RT, Durkin AS, Kolonay JF, Madupu R, Nelson WC, Ayodeji B, Kraul M, Shetty J, Malek J, Van Aken SE, Riedmuller S, Tettelin H, Gill SR, White O, Salzberg SL, Hoover DL, Lindler LE, Halling SM, Boyle SM, Fraser CM: The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc Natl Acad Sci USA 2002;99:13148–13153.

Paulsen IT, Sliwinski MK, Saier MH Jr: Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J Mol Biol 1998;277:573–592.

Pellegrini M: Computational methods for protein function analysis. Curr Opin Chem Biol 2001;5:46–50.

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO: Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci USA 1999;96:4285–4288.

Postma PW, Lengeler JW, Jacobson GR: Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 1993;57:543–594.

Ren Q, Kang KH, Paulsen IT: TransportDB: a relational database of cellular membrane transport systems. Nucleic Acids Res 2004; 32:D284–D288.

Ren Q, Paulsen IT: Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLoS Comput Biol 2005;1:190–201.

Saier MH Jr: Classification of Transmembrane Transport Systems in Living Organisms. San Diego, Academic Press, 1999.

Saier MH Jr: A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev 2000; 64:354–411.

Sweet G, Gandor C, Voegele R, Wittekindt N, Beuerle J, Truniger V, Lin EC, Boos W: Glycerol facilitator of *Escherichia coli*: cloning of *glpF* and identification of the *glpF* product. J Bacteriol 1990;172:424–430.

Tolner B, Ubbink-Kok T, Poolman B, Konings WN: Cation-selectivity of the L-glutamate transporters of *Escherichia coli*, *Bacillus stearothermophilus* and *Bacillus caldotenax*: dependence on the environment in which the proteins are expressed. Mol Microbiol 1995;18:123–133.

Tucker AM, Winkler HH, Driskell LO, Wood DO: S-adenosylmethionine transport in *Rickettsia prowazekii*. J Bacteriol 2003;185: 3031–3035.

Vandepoele K, Van de Peer Y: Exploring the plant transcriptome through phylogenetic profiling. Plant Physiol 2005;137:31–42.

Viitanen P, Newman MJ, Foster DL, Wilson TH, Kaback HR: Purification, reconstitution, and characterization of the *lac* permease of *Escherichia coli*. Methods Enzymol 1986; 125:429–452.

Wu M, Ren Q, Durkin AS, Daugherty SC, Brinkac LM, Dodson RJ, Madupu R, Sullivan SA, Kolonay JF, Haft DH, Nelson WC, Tallon LJ, Jones KM, Ulrich LE, Gonzalez JM, Zhulin IB, Robb FT, Eisen JA: Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. PLoS Genet 2005;1:e65.

Zientz E, Dandekar T, Gross R: Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol Mol Biol Rev 2004;68:745–770.