Comparative Analyses of Fundamental Differences in Membrane Transport Capabilities in Prokaryotes and Eukaryotes

Qinghu Ren, Ian T. Paulsen[*]

The Institute for Genomic Research, Rockville, Maryland, United States of America

Whole-genome transporter analyses have been conducted on 141 organisms whose complete genome sequences are available. For each organism, the complete set of membrane transport systems was identified with predicted functions, and classified into protein families based on the transporter classification system. Organisms with larger genome sizes generally possessed a relatively greater number of transport systems. In prokaryotes and unicellular eukaryotes, the significant factor in the increase in transporter content with genome size was a greater diversity of transporter types. In contrast, in multicellular eukaryotes, greater number of paralogs in specific transporter families was the more important factor in the increase in transporter content with genome size. Both eukaryotic and prokaryotic intracellular pathogens and endosymbionts exhibited markedly limited transport capabilities. Hierarchical clustering of phylogenetic profiles of transporter families, derived from the presence or absence of a certain transporter family, showed that clustering patterns of organisms were correlated to both their evolutionary history and their overall physiology and lifestyles.

Introduction

Membrane transport systems play essential roles in cellular metabolism and activities. 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 [1]. Various transport systems differ in their putative membrane topology, energy coupling mechanisms, and substrate specificities [2]. Among the prevailing energy sources are adenosine triphosphate (ATP), phosphoenolpyruvate, and chemiosmotic energy in the form of sodium ion or proton electrochemical gradients.

The transporter classification system (http://www.tcdb.org/) represents a systematic approach to classify transport systems according to their mode of transport, energy coupling mechanism, molecular phylogeny, and substrate specificity [2–5]. Transport mode and energy coupling mechanism serve as the primary basis for classification because of their relatively stable characteristics. There are four major classes of solute transporters in the transporter classification system: channels, primary (active) transporters, secondary transporters, and group translocators. Transporters of unknown mechanism or function are included as a distinct class. Channels are energy-independent transporters that transport water, specific types of ions, or hydrophilic small molecules down a concentration or electrical gradient; they have higher rates of transport and lower stereospecificity than the other transporter classes (e.g., *Escherichia coli* GlpF glycerol channel [6]). Primary active transporters (e.g., *Lactococcus lactis* LmrP multidrug efflux pump [7]) couple the transport process to a primary source of energy (ATP hydrolysis). Secondary transporters utilize an ion or solute electrochemical gradient, e.g., proton/sodium motive force, to drive the transport process. *E. coli* LacY lactose permease [8,9] is probably one of the best characterized secondary transporters [10]. Group translocators modify their substrates during the transport process. For example, *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 [11,12]. Each transporter class is further classified into individual families and subfamilies according to their function, phylogeny, and/or substrate specificity [3].

Since the advent of genomic sequencing technologies, the complete sequences of over 200 prokaryotic and eukaryotic genomes have been published to date, representing a wide range of species from archaea to human. There are also more than 1,100 additional genome sequencing projects currently underway around the world (Gold Genomes Online Database, http://www.genomesonline.org/) [13,14]. Convenient and effective computational methods are required to handle and analyze the immense amount of data generated by the whole-genome sequencing projects. An in-depth look at transport proteins is vital to the understanding of the metabolic capability of sequenced organisms. However, it is often problematic to annotate these transport proteins by current primary annotation methods because of the occurrence of large and complex transporter gene families, such as the ATP-binding cassette (ABC) superfamily [15,16] and the major facilitator superfamily (MFS) [17,18], and the presence of multiple transporter gene paralogs in many organisms. We

Abbreviations: ABC, adenosine triphosphate–binding cassette; ATP, adenosine triphosphate; GIC, glutamate-gated ion channel; MFS, major facilitator superfamily; ORF, open reading frame; PTS, phosphotransferase system

Editor: Peer Bork, EMBL Heidelberg, Germany

* To whom correspondence should be addressed. E-mail: ipaulsen@tigr.org

Synopsis

Membrane transporters are the cell's equivalent of delivery vehicles, garbage disposals, and communication systems—proteins that negotiate through cell membranes to deliver essential nutrients, eject waste products, and help the cell sense environmental conditions around it. Membrane transport systems play crucial roles in fundamental cellular processes of all organisms. The suite of transporters in any one organism also sheds light on its lifestyle and physiology. Up to now, analysis of membrane transporters has been limited mainly to the examination of transporter genes of individual organisms. But advances in genome sequencing have now made it possible for scientists to compare transport and other essential cellular processes across a range of organisms in all three domains of life.

Ren and Paulsen present the first comprehensive bioinformatic analysis of the predicted membrane transporter content of 141 different prokaryotic and eukaryotic organisms. The scientists developed a new computational application of the phylogenetic profiling approach to cluster together organisms that appear to have similar suites of transporters. For example, a group of obligate intracellular pathogens and endosymbionts possess only limited transporter systems in spite of the massive metabolite fluxes one would expect between the symbionts and their host. This is likely due to the relatively static nature of their intracellular environment. In contrast, a cluster of plant/soil-associated microbes encode a robust array of transporters, reflecting the organisms' versatility as well as their exposure to a wide range of different substrates in their natural environment.

Figure 1. Venn Diagram Showing the Distribution of Transporter Families across the Three Domains of Life

DOI: 10.1371/journal.pcbi.0010027.g001

have been working on a systematic genome-wide analysis of cellular membrane transport systems. Previously, we reported a comprehensive analysis of the transport systems in 18 prokaryotic organisms [19,20] and in yeast [21]. Here we expand our analyses to 141 species and compare the fundamental differences in membrane transport systems in prokaryotes and eukaryotes. Phylogenetic profiling of transporter families and predicted substrates was utilized to investigate the relevance of transport capabilities to the overall physiology of prokaryotes and eukaryotes.

Results/Discussion

Numbers of Recognized Transporter Families and Proteins

A total of 40,678 transport proteins from 141 species (Table S1), including 115 Eubacteria, 17 Archaea, and 9 Eukaryota, were predicted by our analysis pipeline. They were classified into 134 families, including 7 families of primary transporters, 80 families of secondary transporters, 32 channel protein families, 2 phosphotransferase systems (PTSs), and 13 unclassified families. Some of these families are very large superfamilies with numerous members, such as the ABC superfamily and MFS, both of which are widely distributed in Eubacteria, Archaea, and Eukaryota. Some are small families with only a single or a few members. The distribution of transporter families varies significantly across the three domains of life (Figure 1). There are 42 eukaryotic-specific families, mostly ion channel families that exist exclusively in multicellular eukaryotic organisms like *Drosophila melanogaster*, *Arabidopsis thaliana*, and humans. These channels are involved in processes like cell communication, signal transduction, and maintenance of internal homeostasis in a multicellular environment. Most of these families are restricted to a single
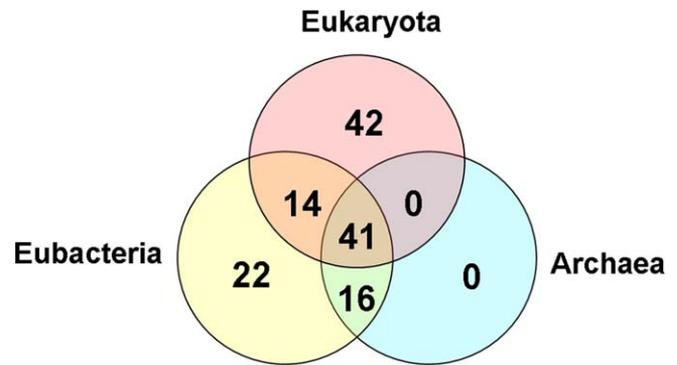
organismal type. Many of them may have arisen later during evolution, after the separation of the three domains. Alternatively, some families may have diverged too extensively from their prokaryotic counterparts to be recognized as homologs. Interestingly, a bacterial homolog to the previously described "eukaryotic-specific" glutamate-gated ion channel (GIC) family of neurotransmitter receptors has now been characterized in *Synechocystis* [22,23], and its orthologs have been identified in other sequenced Cyanobacteria. The *Synechocystis* transporter binds glutamate and forms a K^+-selective ion channel. These observations suggest that eukaryotic GIC family transporters arose from a primordial prokaryotic counterpart.

There are 38 prokaryotic-specific transporter families, of which 22 families exist exclusively in Eubacteria, such as the bacterial sugar PTS systems (see below), and 16 are shared by Eubacteria and Archaea. In contrast to eukaryotic-specific families, which are usually limited to single species, the majority of prokaryotic-specific ones are broadly distributed among prokaryotes. There are no Archaea-specific transporter families currently known. Due to the very limited experimental characterization of Archaea species relative to Eubacteria and Eukaryota, many aspects of the physiology and biochemistry of Archaea are poorly understood [24]. We compared the annotation of membrane proteins in selected species of Archaea and Eubacteria in The Institute for Genomic Research's Comprehensive Microbial Resource database [25]. The percentage of the membrane proteins assigned to the role category of "hypothetical proteins" is significantly greater in Archaea than in Eubacteria (Figure S1). These observations suggest that the sparse functional characterization could be the primary reason for the lack of any known Archaea-specific transporter families.

There are 41 transporter families represented in all three domains of life, highlighting the fundamental importance of these families. These are presumably very ancient families shared by the last common ancestor of Archaea, Eukaryota, and Eubacteria. Most of them were found within the secondary transporter class. These ubiquitous transporter families function in the transport of a diverse spectrum of substrates, including sugars, amino acids, carboxylates, nucleosides, and various cations and anions. There are 14 families shared by Eubacteria and Eukaryota and 16 shared by Eubacteria and Archaea. Some of these families shared only

in two domains may ultimately be discovered in all three domains once a greater diversity of organisms is sequenced.

The overall quantity of recognized transport proteins (Figure 2A) and the percentage relative to the total number of open reading frames (ORFs) (Figure 2B) were compared for the organisms analyzed. Between 2% and 16% of ORFs in prokaryotic and eukaryotic genomes were predicted to encode membrane transport proteins, emphasizing the importance of transporters in the lifestyles of all species. In general, eukaryotic species, especially multicellular eukaryotic organisms, exhibit the largest total number of transport proteins, e.g., *Drosophila* (682 transport proteins, 3.7% of ORFs), *Arabidopsis* (882, 3.5%), *Caenorhabditis elegans* (669, 4.1%), and humans (841, 3.0%). However, the transport proteins of eukaryotic species account for a relatively smaller percentage of total ORFs than in Eubacteria (average 9.3% ± 2.9%) and Archaea (average 6.7% ± 2.3%) species. Considerable variations in the quantity of transport proteins have been observed among species belonging to the same phylogenetic group. For example, α-Proteobacteria species exhibit a wide variety of lifestyles and corresponding differences in transporter content; they range from rhizosphere-dwelling organisms such as *Mesorhizobium loti* and *Sinorhizobium meliloti* [26] with 883 (12.1%) and 826 (13.3%) transport proteins each, to obligate intracellular pathogens or symbionts such as *Rickettsia prowazekii* and *Wolbachia* sp. with 57 (6.8%) and 65 (5.4%) transport proteins, respectively. Overall, prokaryotic obligate endosymbionts and intracellular pathogens, as well as the eukaryotic intracellular parasites (*Plasmodium falciparum* [27] and *Encephalitozoon cuniculi* [28]), possess the most limited repertoire of membrane transporters.

Genome Size versus Diversity of Transporter Families and Numbers of Paralogs

Organisms with a larger genome size and therefore more ORFs generally encode a greater number of transporters [19,29]. In addition to transporters, regulatory genes, secon-
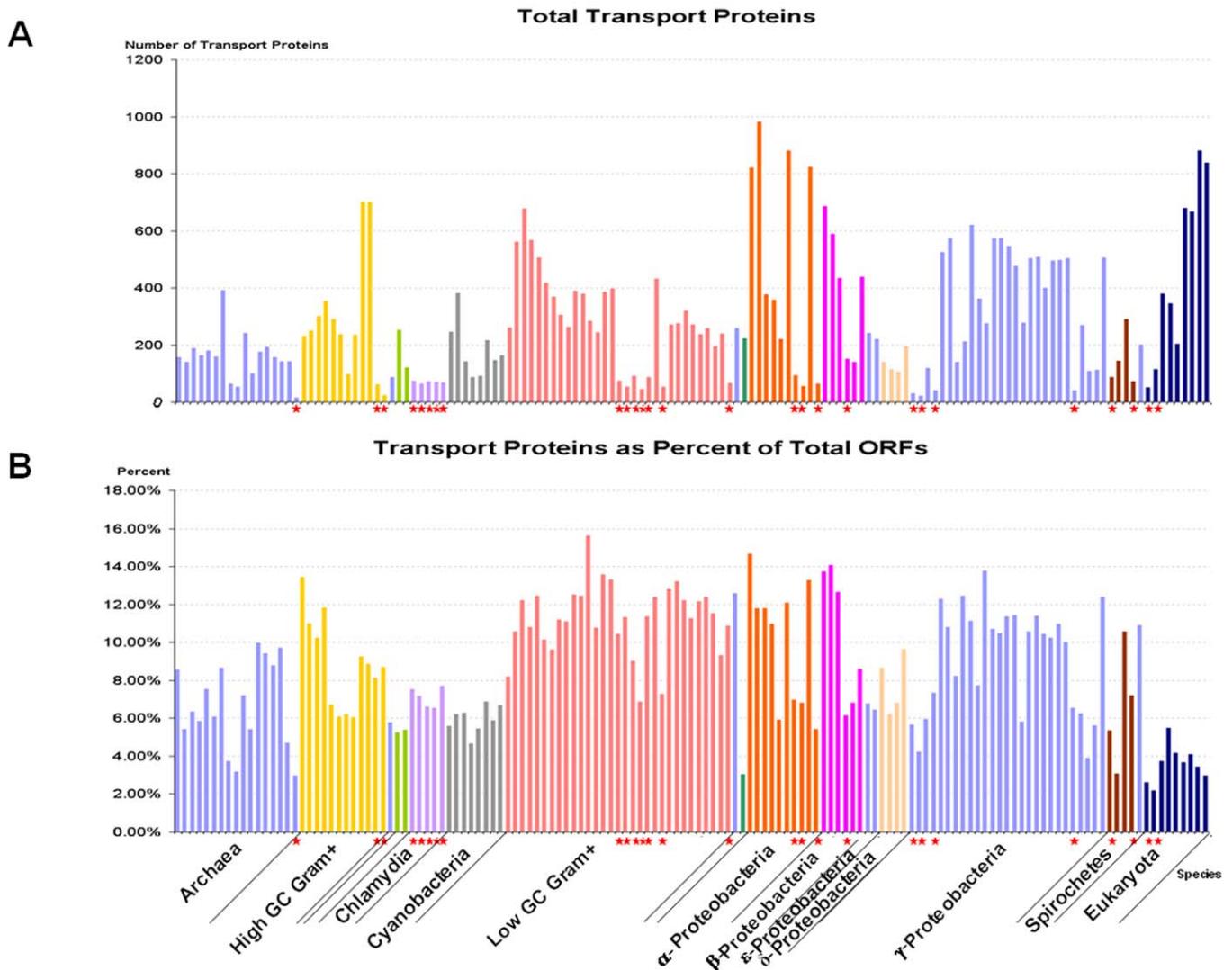

Figure 2. Numbers of Recognized Transport Proteins and Percentage of Total ORFs

The overall numbers of recognized transport proteins (A) and percentage of total ORFs encoding transport proteins (B) were compared for the 141 organisms analyzed. Species from distinct phylogenetic groups are labeled with different colors. The prokaryotic and eukaryotic obligate intracellular parasites/pathogens are marked with red stars.
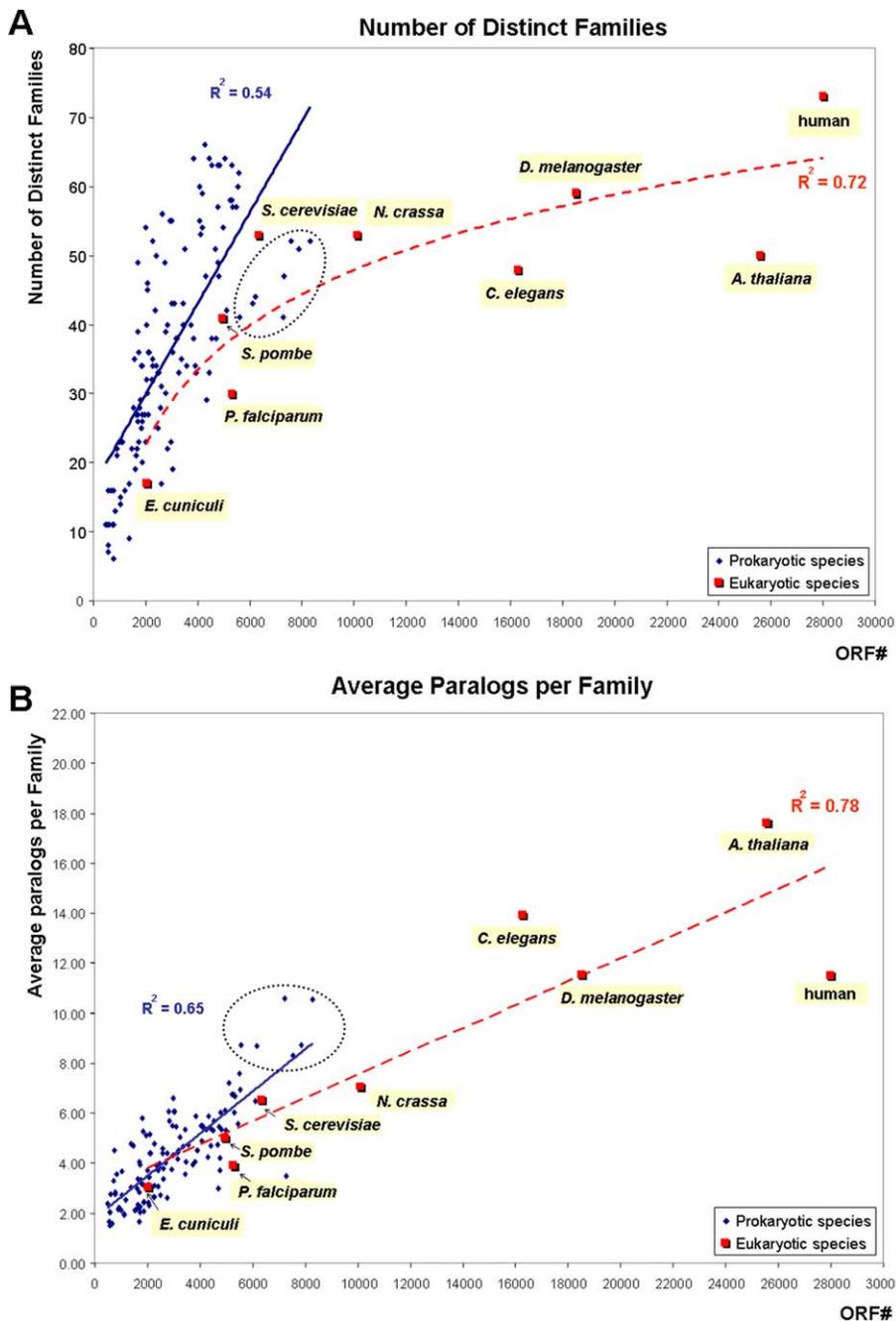
DOI: 10.1371/journal.pcbi.0010027.g002

Figure 3. Number of Total ORFs versus Number of Distinct Transporter Families or Average Number of Paralogs per Family

The number of total ORFs in the genome for each of the 141 sequenced prokaryotic and eukaryotic organisms (x-axis) was plotted as a function of either the number of distinct transporter families (A) or the average number of paralogs per family (B) (y-axis). Blue diamonds represent prokaryotic organisms and red squares represent eukaryotic organisms. Trend line and power correlation R^2 value are shown for prokaryotes and eukaryotes, respectively. A group of α-Proteobacteria are enclosed by a circle (see text for discussion).

dary metabolism genes, and transcription factors, also appear to increase with genome size [29–31]. Two major factors could contribute to the expansion of transporters in organisms with large genome sizes: (1) an increased number of distinct transporter families, and (2) a higher degree of gene duplication or expansion, leading to a greater number of paralogs in certain transporter families. To investigate the relationship between genome size and these two factors, we plotted the total number of ORFs from 141 organisms as a function of either the number of distinct transporter families (Figure 3A), or the average number of paralogs per family (Figure 3B). Prokaryotes and eukaryotes exhibit distinct differences. For prokaryotic species, there is a relatively linear relationship between the genome size and the number of transporter families ($R^2 = 0.54$) or average number of paralogs ($R^2 = 0.65$). As genome size increases, the rate of increase in the number of families per organism is approximately eight times greater than that of the average number

Table 1. The Relative Percentage of Each Transporter Type within Major Phylogenetic Groups

Phylogenetic Group[a]	Primary Transporters	Ion Channels	Secondary Transporters	PTSs	Unclassified
Archaea (17)	39.1% ± 10.4%	6.8% ± 4.2%	52.0% ± 12.1%	NF	2.3% ± 2.0%
Actinobacteria (12)	46.2% ± 10.7%	3.4% ± 1.4%	48.0% ± 9.3%	1.1% ± 1.2%	1.4% ± 1.0%
Chlamydia (5)	39.9% ± 3.8%	NF	49.6% ± 3.1%	8.3% ± 1.1%	2.2% ± 0.1%
Cyanobacteria (8)	51.7% ± 8.0%	8.4% ± 2.5%	38.3% ± 7.6%	NF	1.6% ± 0.5%
Firmicutes (30)	45.6% ± 11.9%	4.2% ± 1.6%	38.7% ± 15.2%	10.9% ± 6.1%	1.1% ± 1.0%
Proteobacteria-alpha (10)	43.9% ± 12.4%	2.9% ± 2.0%	50.0% ± 13.1%	1.5% ± 1.6%	1.6% ± 1.3%
Proteobacteria-beta (6)	41.0% ± 6.8%	3.2% ± 0.7%	51.4% ± 7.5%	2.5% ± 1.3%	2.0% ± 1.8%
Proteobacteria-gamma (27)	30.8% ± 5.6%	5.1% ± 2.5%	55.6% ± 9.3%	7.2% ± 7.2%	1.4% ± 1.0%
Proteobacteria-delta (2)	48.5% ± 8.3%	6.0% ± 0.6%	42.1% ± 5.7%	1.8% ± 2.5%	1.7% ± 0.4%
Proteobacteria-epsilon (3)	34.7% ± 3.4%	4.7% ± 2.5%	57.6% ± 2.5%	NF	3.0% ± 0.8%
Spirochetes (4)	45.3% ± 16.1%	3.1% ± 2.3%	44.2% ± 8.6%	5.9% ± 8.4%	1.6% ± 1.2%
Fungi (3)	15.3% ± 2.5%	4.2% ± 0.3%	77.9% ± 2.2%	NF	1.8% ± 0.2%
Protozoa (1)	48.4%	1.6%	50.0%	NF	0.0%
Microsporidia (1)	41.9%	11.6%	46.5%	NF	0.0%
Nematodes (1)	11.7%	31.1%	56.5%	NF	0.6%
Insects (1)	13.7%	27.9%	56.6%	NF	1.4%
Plants (1)	20.2%	11.9%	65.5%	NF	2.4%
Primates (1)	14.9%	43.3%	38.9%	NF	1.6%

[a]Number of organisms analyzed is indicated in the parenthesis.
NF, not found.
DOI: 10.1371/journal.pcbi.0010027.t001

of paralogs per family. The increase in genome size can only partially explain the expansion of transporter families and paralogs (as indicated by the correlation R^2 value). The strain-specific properties and lifestyles could also have an impact. For example, a group of α-Proteobacteria exhibit the most paralogs per family but have relatively lower diversity of transporter families. These organisms include rhizobial microsymbionts *M. loti, S. meliloti,* and *Bradyrhizobium japonicum* [26], and a closely related plant pathogen, *Agrobacterium tumefaciens* (enclosed by a circle on Figure 3). All of these organisms have more ABC transporters than any other sequenced organisms [29]. ABC family transporters mediate the uptake of a variety of nutrients and the extrusion of drugs and metabolite wastes. Having a large complement of high-affinity ABC uptake systems may be an advantage for organisms in the competition among microbes for nutrients. Two *Streptomyces* species, *St. avermitilis* and *St. coelicolor,* also exhibit a similar trend, with a significant expansion of the ABC and MFS family transporters.

The number of eukaryotic species analyzed is smaller, so it is more difficult to draw robust conclusions. The single-celled eukaryotes such as the yeasts appear to display characteristics similar to those of the prokaryotes, showing expansions in both transporter families and paralogs as genome size increases, with the former being a more important factor. However, in multicellular eukaryotic organisms such as animals and plants, the tremendous number of paralogs in certain transporter families accounts for a significant portion of the increase of transporters. Although multicellular eukaryotes exhibit fewer transporter families than some of the prokaryotic species, they have generated an extraordinary number of paralogs by gene duplication or expansion within certain families, like the ABC superfamily, MFS, and the voltage-gated ion channel superfamily. For example, the *Arabidopsis* genome encodes 110 paralogs of the ABC super-family [32,33] and 92 paralogs of the MFS.

These differences in the relative abundances of transporter

paralogs and distinct transporter families probably represent fundamental differences in transporter needs or priorities of these organisms. Multicellular organisms with many apparently redundant transporter paralogs appear to be utilizing a strategy of specialization. Many of their closely related paralogous transporters are presumably expressed only in specific tissues or subcellular localizations, or at specific developmental time points. Many appear to be involved in cell–cell communication and signal transduction processes, emphasizing the importance of intercellular communication in complex multicellular organisms. In contrast, the single-celled prokaryotes and eukaryotes, with relatively fewer paralogs but a greater emphasis on numbers of different families of transporters, appear to be utilizing a strategy of diversification. This probably reflects that one of the primary roles of membrane transport systems in these organisms is nutrient acquisition. A greater diversity of transporter types presumably allows for a broader range of substrate utilization.

Distribution of Transporter Types According to Energy Coupling Mechanism

A wide range of variations were observed in the relative usage of energy coupling mechanisms to drive transport processes among the prokaryotes and eukaryotes analyzed. Table 1 shows the relative percentage of each transporter type in organisms from major phylogenetic groups. Transporters were categorized into five major types according to transport mode and energy coupling mechanism: primary transporters, secondary transporters, ion channels, group translocators, and unclassified. Primary and secondary carriers are ubiquitous, being present in all organisms analyzed. However, their percentage among the total transporters varies greatly (12%–78% for primary carriers and 17%–80% for secondary carriers). In prokaryotic and unicellular eukaryotic systems, primary and secondary carriers are the predominant types of transporters, together

contributing more than 90% of the total transporters. Channel proteins make up a greater percentage (12%–43%) in higher eukaryotic organisms.

Compared to eukaryotes, prokaryotic organisms rely heavily on primary active transporters, largely because of the usage of ABC uptake systems that are absent in eukaryotes [34]. Organisms with the highest percentage of primary transporters generally belong to one of the three groups. (1) The first group includes organisms that lack a citrate cycle and an electron transfer chain, and therefore can only generate a proton motive force by indirect methods such as substrate-level phosphorylation followed by ATP hydrolysis. These organisms include *Mycoplasma* spp., spirochetes, *Streptococcus* spp., *Tropheryma whipplei, Mycobacterium leprae, Thermoanaerobacter tengcongensis,* and *Thermotoga maritime*. ATP is their primary source of energy, and therefore is most frequently used to drive nutrient uptake and maintain ion homeostasis. (2) The second group includes photosynthetic organisms with the ability to synthesize an ATP pool via photosynthesis, including *Synechocystis* sp., *Nostoc* sp., and *Thermosynechococcus elongates*. (3) The third group is a group of α-Proteobacteria that possess a significant expansion of the ABC superfamily [29], including soil/plant-associated bacteria, such as *M. loti* [26], *S. meliloti* [26], *A. tumefaciens,* and related human/animal pathogens such as *Brucella suis*. Unlike the first two groups, in which the usage of primary transporters seems to be predicated on bioenergetic constraints, the expansion of the ABC transporter family in these α-Proteobacteria does not have any obvious energetic explanations. Instead, it may reflect an organismal requirement for high-affinity transport since ABC transporters typically show higher substrate affinities than most secondary transporters.

The PTS is only present in a subset of Eubacteria, while completely lacking in Archaea and Eukaryota. Gram-negative enteric bacteria, such as *E. coli, Shigella flexneri,* and *Salmonella typhimurium*, as well as Gram-positive species associated with the human gastrointestinal tract, like *Listeria monocytogenes* and *Lactobacillus plantarum,* encode the most abundant PTS systems. Owing to the absorption capacity and efficiency of the intestine, these species have to compete with hundreds of other types of bacteria in an environment containing only small amounts of free carbohydrates or other easily absorbable forms of nutrients. The enrichment of sugar PTS systems in these species could be an advantage to thrive in their ecological niches.

Channel proteins contribute a relatively smaller percentage of transporters in the prokaryotic species we analyzed, and their functions in vivo are largely unknown. Nine organisms lack recognizable channels, including *Chlamydia* spp., *T. whipplei, Treponema pallidum, Wolbachia* sp., and *R. prowazekii,* all of which are obligate intracellular pathogens/symbionts. All other prokaryotic species, including all extremophiles sequenced to date, encode channel proteins, suggesting these channels could function in responding promptly to osmotic and other environmental stresses [35]. Intracellular pathogens and endosymbionts may not need water or ion channels because of their relatively static intracellular environment and may largely depend on their host organisms for maintenance of ion homeostasis.

The percentage of channel proteins increases significantly in multicellular eukaryotes. In animals, these consist largely of ion channels with communication roles, such as in signal transduction, or roles as sensors for external stimuli. For example, members in the ligand-gated ion channel family [36] and the GIC family [37] are activated by major excitatory (glutamate) and inhibitory neurotransmitters (GABA) and participate in neuronal communication in the brain [38]. Recent studies show that some subunits of ligand-gated ion channels and GIC-type channels are expressed prominently during embryonic and postnatal brain development, while others are expressed mainly in the adult brain, suggesting that a switch in subunit composition may be required for normal brain development [38]. In plants, approximately one-third of the channel proteins are aquaporins (water channels) [39], many of which show a cell-specific expression pattern in the root, emphasizing the importance of regulating and maintaining turgor pressure through the plant [40].

Three fungal species, *Saccharomyces cerevisiae, Schizosaccharomyces pombe,* and *Neurospora crassa,* possess the largest portion of secondary transporters (76%–80%), mainly because of the prominent gene expansion of two types of functionally diverse MFS family transporters: (1) drug efflux pumps, which could play roles in the secretion of secondary metabolites, toxic compounds, and signaling molecules, and (2) sugar symporters, which could allow a broader range of sugar utilization [41,42].

Phylogenetic Profiling of Transporter Family and Substrate Shows Strong Correlations to Organisms' Overall Physiology

The phylogenetic profile of a given protein is a string that encodes the presence or absence of that protein in every fully sequenced genome. Proteins that function together in a pathway or a common structural complex are likely to evolve in a correlated fashion, and therefore tend to be either preserved or eliminated together in a new species during evolution [43,44]. Phylogenetic profiling has been an effective way to detect conserved core genes, species-specific gene families, lineage-specific gene family expansions [45], and subcellular localization of proteins [46]. It can also facilitate the prediction of physical and functional interactions and assist in the deduction of the functions of genes that have no well-characterized homologs [47,48].

We have undertaken a novel application of phylogenetic profiling to investigate the presence or absence of transporter protein families across sequenced genomes. To our knowledge this represents the first application of a phylogenetic profiling approach using protein families rather than individual proteins as the unit of comparison. With the data on membrane transport systems from 141 fully sequenced organisms, we were able to construct the phylogenetic profiles for each transporter family (Figures 4 and S2). Hierarchical clustering of phylogenetic profiles showed a strong correlation between the observed clustering pattern and phylogeny, with Eubacteria, Archaea, and Eukaryota clearly separated into different clusters. Inside the bacterial cluster, Gram-positive bacteria, Proteobacteria, *Chlamydia,* and Cyanobacteria are also clearly defined into different groups. Given that the profiling approach solely utilizes presence or absence of a transporter family and does not use sequence similarity directly, this indicates that the types of transporters utilized by organisms are related to their evolutionary history. Additionally, the clustering appears to be influenced by habitat or lifestyle of organisms. For
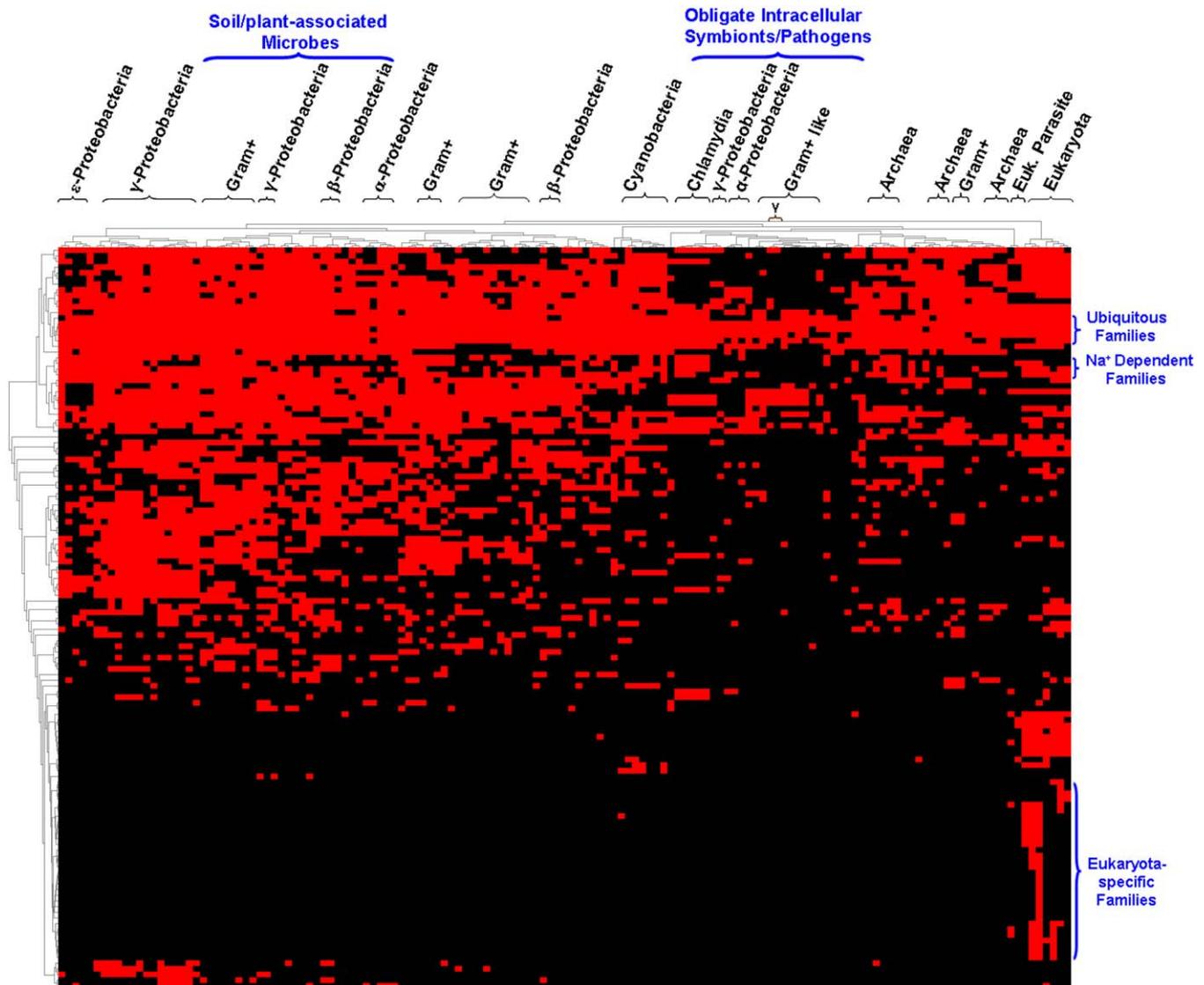
Figure 4. Phylogenetic Profiling of Transporter Families

Phylogenetic profiles were created for each transporter family. Each profile is a string with 141 entries (number of organisms analyzed). If a given family is present in an organism, the value one is assigned at this position (red). If not, zero is assigned (black). Organisms and transporter families were clustered according to the similarity of their phylogenetic profiles.

DOI: 10.1371/journal.pcbi.0010027.g004

example, the obligate intracellular pathogens/symbionts and a collection of soil/plant-associated microbes are separated into two distinct superclusters (Figure 5).

The obligate intracellular pathogens/symbionts cluster includes a group of phylogenetically diverse organisms, including *Chlamydia* spp. (pathogens); γ-Proteobacteria such as *Buchnera* spp., *Wigglesworthia glossinidia brevipalpis,* and *Candidatus Blochmannia floridanus* (endosymbionts); α-Proteobacteria such as *Wolbachia* sp. (endosymbiont) and *R. prowazekii* (pathogen); Gram-positive-like organisms *Mycoplasma* spp. and *T. whipplei* (pathogens); Spirochetes such as *Tr. pallidum* and *Borrelia burgdorferi* (pathogens); and an archaeal symbiont, *Nanoarchaeum equitans.* Organisms in this cluster share an obligate intracellular lifestyle as well as reduced genome size. The clustering does not appear to be due to genome size alone as nonobligate intracellular organisms with small genome sizes do not fall into this cluster. One

possibility is that the transport needs of these obligate intracellular organisms are more specialized than those of environmental organisms because of the much more static nature of their intracellular environments. This may have allowed them to shed, for example, transporters for alternative nitrogen/carbon sources, osmoregulatory functions, and ion homeostasis. Similar to their prokaryotic counterparts, two eukaryotic intracellular parasites, *P. falciparum* and *En. cuniculi,* form a distinct cluster separate from the other eukaryotes.

The soil/plant-associated microbe cluster also contains species from various phylogenetic groups, such as Actinobacteria *(Corynebacterium* and *Streptomyces),* Firmicutes *(Bacillus* and *Oceanobacillus),* α-Proteobacteria *(Brucella, Agrobacterium, Mesorhizobium, Sinorhizobium,* and *Bradyrhizobium),* β-Proteobacteria *(Bordetella* and *Ralstonia),* γ-Proteobacteria *(Pseudomonas* and *Rhodopseudomonas),* δ-Proteobacteria *(Geobacter),* Deino-
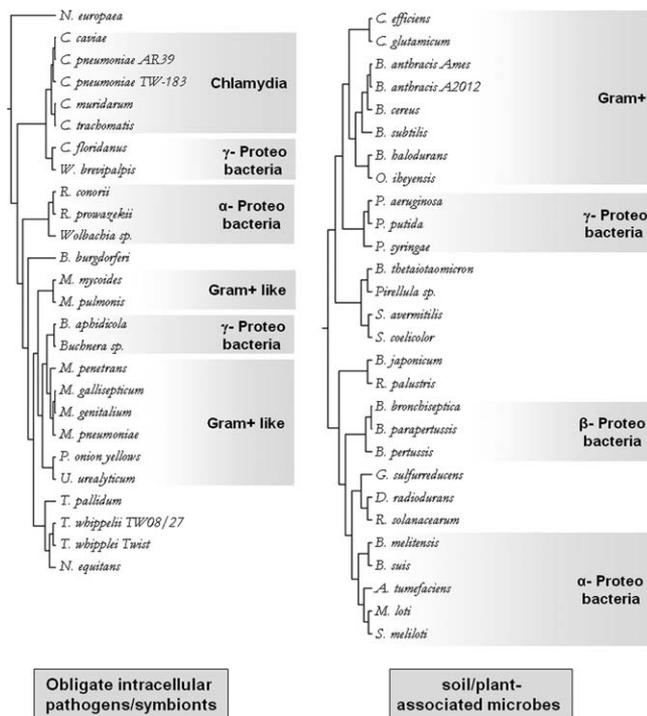
Figure 5. Hierarchial Clustering of Phylogenetic Profiles of Obligate Intracellular Pathogens/Symbionts versus Soil/Plant-Associated Microbes

Detailed view of two clusters of organisms generated by hierarchical clustering of their phylogenetic profiles of transporter families: obligate intracellular pathogens/symbionts and soil/plant-associated microbes.

DOI: 10.1371/journal.pcbi.0010027.g005

coccus (*Deinococcus radiodurans*), Planctomycetes (*Pirellula* sp.), and Bacteroidetes (*Bacteroides thetaiotaomicron*). All of these organisms possess a robust collection of transporter systems. It is unlikely that these species are merely clustered by their genome sizes because some species in this cluster have relatively smaller genome sizes, like *Corynebacterium efficiens* (3.0 Mb), *D. radiodurans* (3.2 Mb), and *Brucella melitensis* (3.2 Mb). In addition, hierarchical clustering of organisms exclusively by genome size generates clusters with no apparent phylogenetic relationship (data not shown). The similarity of phylogenetic profiles of organisms in this cluster 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 pathogens, e.g., *Bordetella*, *Brucella*, *Bacillus anthracis* [26], and *Bacteroides thetaiotaomicron*, are also grouped in this cluster. All of these pathogens have close relatives that are soil- or plant-associated environmental organisms [49–52], so their transport capabilities probably reflect a combination of their evolutionary heritage, original environmental niche, and current transport needs.

To compare the transport capabilities of organisms in the intracellular pathogen/symbiont cluster and the soil/plant-associated microbe cluster, we carried out statistical analysis on their number of transporters, percentage of ORFs encoding transport proteins, and compositions in each transporter type (data not shown). Organisms in the soil/plant-associated microbe cluster on average have about eight times as many transporters as those in the intracellular organism cluster ($p < 0.0001$; p-value denotes the confidence level that the correlation observed is significantly different from the null hypothesis). The difference in the relative percentage of ORFs that are transporters is smaller but still significant (1.5-fold increase, $p < 0.0001$), suggesting that systematic gene loss and genome compaction is one of the important factors in reducing the number of transport proteins in intracellular organisms. 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, a glutamate transporter is encoded in two obligate endosymbionts: the GltP glutamate:proton symporter (DAACS family) [53] in *Candidatus Blochmannia floridanus*, and GltJKL ABC transporter [54] in *Wigglesworthia glossinidia brevipalpis*. These organisms have a truncated citrate cycle that begins with α-ketoglutarate and ends with oxaloactetate [55]. Their citrate cycle could be closed by the transamination of the imported glutamate to aspartate, catalyzed by an aspartate aminotransferase (AspC) that uses oxaloactetate as a cosubstrate and produces α-ketoglutarate. As to the distribution of transporter types, there is no significant difference between these two clusters although intracellular organisms show a higher degree of variation in each transporter type than the plant/soil-associated microbes. 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 determines the set of transporters that the organism cannot afford to lose.

In addition to investigating the relationship between organisms based on their transporter profiles, we also examined the clustering of transporter families. The essentially ubiquitous families, like ABC, MFS, P(F)-type ATPase, that are present in virtually every organism we analyzed, are clustered together. Eukaryotic-specific families, most of which are single-organism-specific ion channels, are grouped together. Interestingly, the sodium-ion-dependent families, like neurotransmitter:sodium symporter, alanine/glycine:cation symporter, solute:sodium symporter, and divalent anion:sodium symporter [56–58], are clustered together. Transporters in these families are all symporters that utilize the sodium ion gradient to transport amino acid, solute, and/or divalent ions into cytoplasm. This clustering may suggest that these families co-occur in a specific set of organisms, presumably those reliant on sodium-ion-driven transport.

Previous studies have shown that transporters with similar functions characteristically cluster together in phylogenetic analyses; hence, substrate specificity appears to be a conserved evolutionary trait in transporters [19,20,59,60]. The phylogenetic profiles of predicted substrates for all 141 organisms were generated and clustered by MeV (see Figure S3). Overall, similar patterns were observed as with the clustering by families. Organisms were grouped together either by their phylogenetic history or by their physiology or living habits. Ubiquitous substrates (e.g., cation, amino acid, sugar, and phosphate) and eukaryotic-specific substrates (e.g.,

cholesterol, UDP-sugars, and phosphoenolpyruvate) each form distinct clusters.

Distribution of Transporter Families among Species in the Same Genus

With the transporter data from a great diversity of sequenced organisms, we were able to compare the distribution of transporter families in closely related species (i.e., from the same genus) (Figures 6 and S4). In most of the cases we studied, species from the same genus share highly parallel distributions of transporter families. For example, three *Pseudomonas* species, *Ps. aeruginosa* [61], *Ps. putida* [62] and *Ps. syringae* [63], all of which are metabolically versatile soil/plant-associated bacteria, show highly similar patterns of transporter family distribution. Among the 66 transporter families present in this genus, 47 are shared by all three species and 14 are shared by two species (Figure 6A). All three species encode transporters for a diverse spectrum of substrates, including sugars, amino acids, peptides, carboxylates, and various cations and anions.

The distribution of transporter families in three *Corynebacterium* species represents an exception. *Co. glutamicum* [64] and *Co. efficiens* [65] are widely used in the industrial production of amino acids like glutamic acid and lysine by fermentation. The closely related *Co. diphtheriae* [66], however, is a human pathogen causing the respiratory illness diphtheria and lacks amino acid productivity. Compared to the other two species, *Co. diphtheriae* shows a dramatically different transporter family profile (Figure 6B). There are eight families specific to *Co. diphtheriae,* while only one for *Co. glutamicum* and three for *Co. efficiens.* More importantly, *Co. diphtheriae* uses totally different mechanisms to transport potassium ion and C4-dicarboxylates than the other two species. In *Co. diphtheriae*, potassium ions are transported into cytoplasm via a Trk family K^+:H^+ symporter [67], while both *Co. glutamicum* and *Co. efficiens* encode a KUP family potassium ion uptake permease [68]. *Co. diphtheriae* utilizes the DcuABC antiporter system [69] for the uptake of C4-dicarboxylate, while the other species use the ATP-independent tripartite periplasmic symporter systems (TRAP-T family) [70]. The common orthologs of transporters in families specific to one or two *Corynebacterium* species were identified in sequenced high-GC Gram-positive bacteria, and the phylogenetic trees were constructed by the neighbor-joining method (data not shown). For those families with orthologs in *Co. glutamicum* and *Co. efficiens* but not in *Co. diphtheriae,* orthologs were also identified in the majority of high-GC Gram-positive species. The trees of transport protein are similar to the 16S rRNA tree, suggesting certain transporter families in *Co. efficiens* are missing because of specific gene losses. By contrast, *Co. diphtheriae*–specific transporter families, like Dcu, DcuC, and Trk families, tend to have either no apparent orthologs or only distantly related homologs in other sequenced high-GC Gram-positive species, suggesting possible evolutionary gene acquisition events in *Co. diphtheriae.* The recent finding that both gene loss and horizontal gene transfer are responsible for the functional differentiation in amino acid biosynthesis of the three *Corynebacterium* species [71] further supports this conclusion.

All three *Corynebacterium* species share 41 transporter families. Interestingly, although *Co. diphtheriae* shows no amino acid productivity and has a reduced genome size
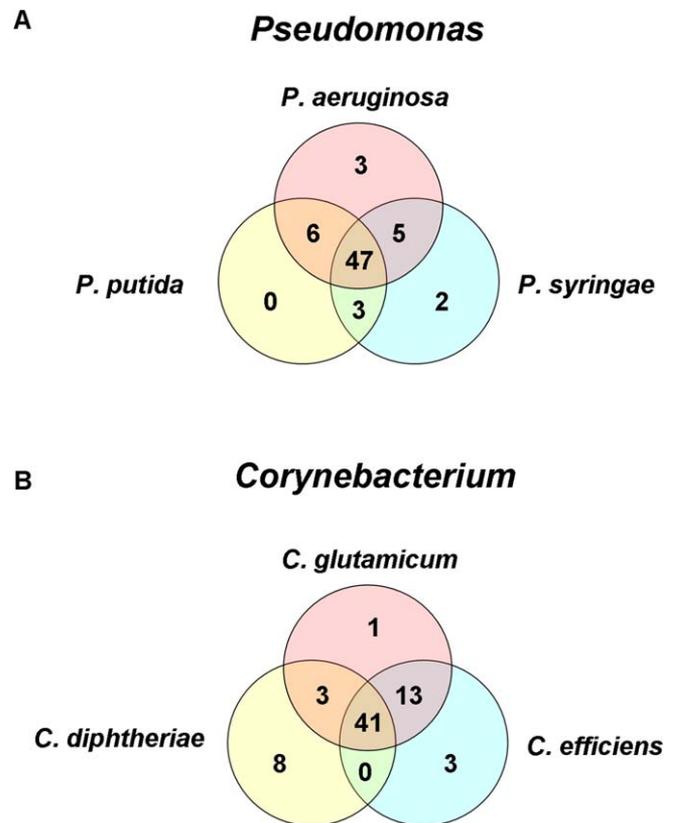
Figure 6. Venn Diagrams Showing the Distribution of Transporter Families among Species Belonging to the Same Genus
(A) Transporter family distribution among three *Pseudomonas* species.
(B) Transporter family distribution among three *Corynebacterium* species.
DOI: 10.1371/journal.pcbi.0010027.g006

[71], all the major types of amino acid exporters in *Co. glutamicum* [72] are conserved in *Co. diphtheriae,* e.g., the LysE family transporter for the export of basic amino acids, the RhtB family transporter for threonine efflux, the ThrE family transporter for threonine and serine export, and the LIV-E family transporter (BrnFE in *Co. glutamicum*), which is a two-component efflux pump exporting branched-chain amino acids [73]. The only difference observed among these organisms is the number of paralogs in the RhtB family: three in *Co. glutamicum,* two in *Co. efficiens,* and only one in *Co. diphtheriae.* The phylogenetic tree of the RhtB family suggests that gene duplication took place in the common ancestor of *Corynebacterium,* and that specific gene loss was responsible for the single RhtB transporter in *Co. diphtheriae.*

Conclusion

The rapid expansion of complete genome sequencing enabled us to conduct analyses of transporter capabilities on the whole-genome level. By comparing the membrane transport systems in Eubacteria, Archaea, and Eukaryota, we could draw conclusions as follows. (1) Eukaryotic species generally encode a larger number of transporters, but transporters account for a smaller percentage of total ORFs in eukaryotic than in prokaryotic species. Prokaryotic obligate intracellular pathogens and endosymbionts, as well as the eukaryotic parasites, possess the most limited repertoire of membrane transporters. (2) Organisms with a larger

genome size tend to have a higher number of transporters. In prokaryotes and unicellular eukaryotes, this increase is primarily due to increased diversity of types of transporter. In multicellular eukaryotes, this increase is largely due to the greater number of paralogs by gene duplication or expansion in certain transporter families. (3) The distribution of different transporter types according to transport mode and energy coupling mechanism generally correlates with organisms' primary mechanism of energy generation. Compared to eukaryotes, prokaryotic species rely heavily on primary (active) transporters. Primary type transporters in Eubacteria and Archaea account for a much larger percentage of total transporters than any other transporter type. This phenomenon may be related to the absence of ABC-type uptake permeases in eukaryotes and, in some cases, the bioenergetic requirements and environmental constraints of prokaryotic organisms. (4) Energy-independent channel proteins are far more numerous in multicellular organisms and are often involved in cell–cell communication and signal transduction processes. Many channels are restricted to a single organismal type. The expression of different subunits of a channel in a timely fashion may be an essential step during embryonic development in mammals. (5) The PTS is only present in a subset of Eubacteria, and is completely absent in Archaea and Eukaryota. The expansion of sugar PTS systems in species dwelling in the gastrointestinal tract could provide the advantage to thrive in their ecological niches. (6) Hierarchical clustering of the phylogenetic profiles of transporter families showed that the distribution of transporter families appears to reflect a combination of evolutionary history and environment and lifestyle factors. (7) The distribution pattern of transporter families in species belonging to the same genus is usually parallel, with some notable exceptions that may reflect specific environmental differences.

Materials and Methods

We developed a semi-automated pipeline to annotate transport systems genome-wide, input the data into TransportDB database, and visualize the result through a Web interface [74]. The complete protein sequences from specific organisms were first searched against our curated database of transport proteins for similarity to known or putative transport proteins using BLAST [75,76]. All of the query proteins with significant hits (E-value < 0.001) were collected and searched against the NCBI nonredundant protein database and Pfam database [77]. Transmembrane protein topology was predicted by TMHMM [78]. A Web-based interface was created to facilitate the annotation processes, which incorporates number of hits to the transporter database, BLAST and HMM search E-value and score, number of predicted transmembrane segments, and the description of top hits to the nonredundant protein database. We also set up direct links between transporter classification family and COG classification [79] so that COG-based searches can inform the

transporter annotation. The results can be viewed at the TransportDB Web site (http://www.membranetransport.org/).

To analyze the phylogenetic profiles of transporter families and predicted substrates, we assigned a profile to each transporter family or substrate. Each profile is a string with 141 entries (number of species analyzed). If a given family is present or a given substrate is transported in certain species, the value one was assigned at these positions (red for transporter families/purple for predicted substrates). If not, zero was assigned (black). Transporter families or substrates were clustered according to the similarity of their phylogenetic profiles using The Institute for Genomic Research's microarray multi-experiment viewer (MeV) [80] with two-dimensional hierarchical clustering as described by Eisen et al. [81].

Supporting Information

Acknowledgments

References

1. Saier MH Jr (1999) Classification of transmembrane transport systems in living organisms. In: VanWinkle L, editor. Biomembrane transport. San Diego: Academic Press. pp. 265–276
2. Saier MH Jr (2000) A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev 64: 354–411.
3. Saier MH Jr (1999) A functional-phylogenetic system for the classification of transport proteins. J Cell Biochem 75 (Suppl 32): 84–94.
4. Busch W, Saier MH Jr (2002) The transporter classification (TC) system, 2002. Crit Rev Biochem Mol Biol 37: 287–337.
5. Busch W, Saier MH (2004) The IUBMB-endorsed transporter classification system. Mol Biotechnol 27: 253–262.
6. Sweet G, Gandor C, Voegele R, Wittekindt N, Beuerle J, et al. (1990) Glycerol facilitator of *Escherichia coli*: Cloning of glpF and identification of the glpF product. J Bacteriol 172: 424–430.
7. Bolhuis H, Poelarends G, van Veen HW, Poolman B, Driessen AJ, et al. (1995) The lactococcal lmrP gene encodes a proton motive force-dependent drug transporter. J Biol Chem 270: 26092–26098.
8. Newman MJ, Foster DL, Wilson TH, Kaback HR (1981) Purification and reconstitution of functional lactose carrier from *Escherichia coli*. J Biol Chem 256: 11804–11808.
9. Viitanen P, Newman MJ, Foster DL, Wilson TH, Kaback HR (1986) Purification, reconstitution, and characterization of the lac permease of *Escherichia coli*. Methods Enzymol 125: 429–452.
10. Abramson J, Smirnova I, Kasho V, Verner G, Iwata S, et al. (2003) The

lactose permease of *Escherichia coli:* Overall structure, the sugar-binding site and the alternating access model for transport. FEBS Lett 555: 96–101.

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

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

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

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

15. Davidson AL, Chen J (2004) ATP-binding cassette transporters in bacteria. Annu Rev Biochem 73: 241–268.

16. Schneider E, Hunke S (1998) ATP-binding cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. FEMS Microbiol Lett 22: 1–20.

17. Marger MD, Saier MH Jr (1993) A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem Sci 18: 13–20.

18. Saier MH Jr, Beatty JT, Goffeau A, Harley KT, Heijne WH, et al. (1999) The major facilitator superfamily. J Mol Microbiol Biotechnol 1: 257–279.

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

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

21. Paulsen IT, Sliwinski MK, Nelissen B, Goffeau A, Saier MH Jr (1998) Unified inventory of established and putative transporters encoded within the complete genome of *Saccharomyces cerevisiae*. FEBS Lett 430: 116–125.

22. Chen GQ, Cui C, Mayer ML, Gouaux E (1999) Functional characterization of a potassium-selective prokaryotic glutamate receptor. Nature 402: 817–821.

23. Mayer ML, Olson R, Gouaux E (2001) Mechanisms for ligand binding to GluR0 ion channels: Crystal structures of the glutamate and serine complexes and a closed apo state. J Mol Biol 311: 815–836.

24. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273: 1058–1073.

25. Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O (2001) The comprehensive microbial resource. Nucleic Acids Res 29: 123–125.

26. Ruder K, Winstead ER, Gibbs MS (2004) A quick guide to sequenced genomes. Rockville (Maryland): Genome News Network. Available: http://www.genomenewsnetwork.org/resources/sequenced_genomes/genome_guide_p1.shtml. Accessed 13 July 2005.

27. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498–511.

28. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450453.

29. Konstantinidis KT, Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A 101: 3160–3165.

30. Cases I, de Lorenzo V, Ouzounis CA (2003) Transcription regulation and environmental adaptation in bacteria. Trends Microbiol 11: 248.

31. Jordan IK, Makarova KS, Spouge JL, Wolf YI, Koonin EV (2001) Lineage-specific gene expansions in bacterial and archaeal genomes. Genome Res 11: 555–565.

32. Sanchez-Fernandez R, Davies TG, Coleman JO, Rea PA (2001) The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. J Biol Chem 276: 30231–30244.

33. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

34. Dean M, Allikmets R (2001) Complete characterization of the human ABC gene family. J Bioenerg Biomembr 33: 475–479.

35. Kung C, Blount P (2004) Channels in microbes: So many holes to fill. Mol Microbiol 53: 373–380.

36. Hong X (1998) Identification of major phylogenetic branches of inhibitory ligand-gated channel receptors. J Mol Evol 47: 323.

37. Nakanishi S, Masu M (1994) Molecular diversity and functions of glutamate receptors. Annu Rev Biophys Biomol Struct 23: 319–348.

38. Lujan R, Shigemoto R, Lopez-Bendito G (2005) Glutamate and GABA receptor signalling in the developing brain. Neuroscience 130: 567–580.

39. Johanson U, Karlsson M, Johansson I, Gustavsson S, Sjovall S, et al. (2001) The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. Plant Physiol 126: 1358–1369.

40. Javot H, Maurel C (2002) The role of aquaporins in root water uptake. Ann Bot 90: 301–313.

41. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422: 859–868.

42. Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, et al. (2004) Lessons from the genome sequence of *Neurospora crassa*: Tracing the path from genomic blueprint to multicellular organism. Microbiol Mol Biol Rev 68: 1–108.

43. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc Natl Acad Sci U S A 96: 4285–4288.

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

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

46. Marcotte EM, Xenarios I, van der Bliek AM, Eisenberg D (2000) Localizing proteins in the cell from their phylogenetic profiles. Proc Natl Acad Sci U S A 97: 12115–12120.

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

48. Kriventseva EV, Biswas M, Apweiler R (2001) Clustering and analysis of protein families. Curr Opin Struct Biol 11: 334–339.

49. Paulsen IT, Seshadri R, Nelson KE, Eisen JA, Heidelberg JF, et al. (2002) The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc Natl Acad Sci U S A 99: 13148–13153.

50. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, et al. (2003) A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. Science 299: 2074–2076.

51. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, et al. (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet 35: 32–40.

52. Gerlach G, von Wintzingerode F, Middendorf B, Gross R (2001) Evolutionary trends in the genus *Bordetella*. Microbes Infect 3: 61–72.

53. Tolner B, Ubbink-Kok T, Poolman B, Konings WN (1995) 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 18: 123–133.

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

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

56. Reizer J, Reizer A, Saier MH Jr (1994) A functional superfamily of sodium/solute symporters. Biochim Biophys Acta 1197: 133–166.

57. Saier MH Jr, Eng BH, Fard S, Garg J, Haggerty DA, et al. (1999) Phylogenetic characterization of novel transport protein families revealed by genome analyses. Biochim Biophys Acta 1422: 1–56.

58. Pajor AM (2000) Molecular properties of sodium/dicarboxylate cotransporters. J Membr Biol 175: 1–8.

59. Pao SS, Paulsen IT, Saier MH Jr (1998) Major facilitator superfamily. Microbiol Mol Biol Rev 62: 1–34.

60. Jack DL, Paulsen IT, Saier MH (2000) The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. Microbiology 146: 1797–1814.

61. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, et al. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406: 959–964.

62. Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol 4: 799–808.

63. Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, et al. (2003) The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. Proc Natl Acad Sci U S A 100: 10181–10186.

64. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, et al. (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of aspartate-derived amino acids and vitamins. J Biotechnol 104: 5–25.

65. Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, et al. (2003) Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. Genome Res 13: 1572–1579.

66. Cerdeno-Tarraga AM, Efstratiou A, Dover LG, Holden MTG, Pallen M, et al. (2003) The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. Nucleic Acids Res 31: 6516–6523.

67. Schlosser A, Meldorf M, Stumpe S, Bakker EP, Epstein W (1995) TrkH and its homolog, TrkG, determine the specificity and kinetics of cation transport by the Trk system of *Escherichia coli*. J Bacteriol 177: 1908–1910.

68. Trchounian A, Kobayashi H (1999) Kup is the major K+ uptake system in *Escherichia coli* upon hyper-osmotic stress at a low pH. FEBS Lett 447: 144–148.

69. Engel P, Kramer R, Unden G (1994) Transport of C4-dicarboxylates by anaerobically grown *Escherichia coli*. Energetics and mechanism of exchange, uptake and efflux. Eur J Biochem 222: 605–614.

70. Kelly DJ, Thomas GH (2001) The tripartite ATP-independent periplasmic

(TRAP) transporters of bacteria and archaea. FEMS Microbiol Rev 25: 405–424.

71. Nishio Y, Nakamura Y, Usuda Y, Sugimoto S, Matsui K, et al. (2004) Evolutionary process of amino acid biosynthesis in *Corynebacterium* at the whole genome level. Mol Biol Evol 21: 1683–1691.

72. Eggeling L, Sahm H (2003) New ubiquitous translocators: Amino acid export by *Corynebacterium glutamicum* and *Escherichia coli.* Arch Microbiol 180: 155–160.

73. Kennerknecht N, Sahm H, Yen MR, Patek M, Saier MH Jr, et al. (2002) Export of L-isoleucine from *Corynebacterium glutamicum:* A two-gene-encoded member of a new translocator family. J Bacteriol 184: 3947–3956.

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

75. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

76. Altschul SF, Gish W (1996) Local alignment statistics. Methods Enzymol 266: 460–480.

77. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R (1998) Pfam: Multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res 26: 320–322.

78. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305: 567–580.

79. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, et al. (2001) The COG database: New developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29: 22–28.

80. Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. Biotechniques 34: 374–378.

81. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.