ELSEVIER

Biochimica et Biophysica Acta 1562 (2002) 6–31

BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Review

Protein-translocating outer membrane porins of Gram-negative bacteria

Ming-Ren Yen[a,1], Christopher R. Peabody[a,2], Salar M. Partovi[a], Yufeng Zhai[a], Yi-Hsiung Tseng[b], Milton H. Saier Jr.[a,*]

[a]*Division of Biology 0116, 9500 Gilman Drive, University of California at San Diego, La Jolla, CA 92093-0116, USA*
[b]*Institute of Molecular Biology, National Chung Hsing University, Taichung 402, Taiwan, ROC*

Received 16 October 2001; received in revised form 17 January 2002; accepted 17 January 2002

Abstract

Five families of outer membrane porins that function in protein secretion in Gram-negative bacteria are currently recognized. In this report, these five porin families are analyzed from structural and phylogenetic standpoints. They are the fimbrial usher protein (FUP), outer membrane factor (OMF), autotransporter (AT), two-partner secretion (TPS) and outer membrane secretin (Secretin) families. All members of these families in the current databases were identified, and all full-length homologues were multiply aligned for structural and phylogenetic analyses. The organismal distribution of homologues in each family proved to be unique with some families being restricted to proteobacteria and others being widespread in other bacterial kingdoms as well as eukaryotes. The compositions of and size differences between subfamilies provide evidence for specific orthologous relationships, which agree with available functional information and intra-subfamily phylogeny. The results reveal that horizontal transfer of genes encoding these proteins between phylogenetically distant organisms has been exceptionally rare although transfer within select bacterial kingdoms may have occurred. The resultant in silico analyses are correlated with available experimental evidence to formulate models relevant to the structures and evolutionary origins of these proteins. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Phylogeny; Evolution; Gram-negative bacteria; Outer membrane; Porins; β-barrel

1. Introduction

The dual membrane envelopes of Gram-negative bacteria provide two barriers of unlike nature that pose formidable problems concerning the transport of molecules into and out of these diverse organisms. While nutrients and essential cofactors must be actively transported into the cells, end products of metabolism, toxic substances and secreted macromolecules must be actively extruded. Specific transport systems have evolved to achieve these goals. The diversity of such systems currently recognized in Gram-negative bacteria far exceeds that recognized in Gram-positive bacteria, archaea or eukarya [1].

Protein secretion proves to be illustrative of this fact. Thus, while Gram-positive bacteria, eukaryotes and archaea exhibit just three known, functionally characterized protein secretory systems for transport across cytoplasmic/endoplasmic reticular membranes, first, the so-called general secretory pathway (GSP) or type II secretory pathway (IISP) [2] (T.B. Cao and M.H. Saier, Jr., submitted); second, the cytochrome oxidase biogenesis (Oxa1/YidC) pathway [3–5]; and third, the twin arginine targeting/translocation (Tat) pathway, Gram-negative bacteria have multiple such systems for protein transport across their cytoplasmic membranes as well as multiple systems for transport across their outer, lipopolysaccharide-containing membranes [6]. Moreover, distinct pathways may overlap since some evidence suggests that the Tat pathway can feed into the GSP [7], and insertion of integral membrane proteins via the GSP may sometimes function in conjunction with the Oxa1/YidC system (see Refs. [4,5] for reviews).

Among the outer membrane protein (OMP) secreting porins, those of types II, III and IVSP have recently been characterized from phylogenetic and structural standpoints [2,8–11] (T.B. Cao and M.H. Saier, Jr., submitted). While types IISP and IIISP systems use oligomeric secretins [10], type IVSP systems may use heterooligomeric structures consisting of three sequence dissimilar proteins [8]. Four additional protein secreting OMPs are currently recognized

---

[*] Corresponding author. Tel.: +1-858-534-4084; fax: +1-858-534-7108. *E-mail address:* msaier@ucsd.edu (M.H. Saier Jr.).

[1] Present address: Institute of Molecular Biology, National Chung Hsing University, Taichung 402, Taiwan, ROC.

[2] Present address: Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720, USA.

PII: S0005-2736(02)00359-0

in Gram-negative bacteria. Three of these four families have been analyzed previously, the fimbrial usher protein (FUP) family [12–14], the outer membrane factor (OMF) family [15–17] and the autotransporter (AT) family [18]. The fourth family, the two-partner secretion (TPS) family, has not, to our knowledge, been carefully examined from a phylogenetic standpoint [19]. In this communication, we provide updates of the five families of outer membrane porins that are believed to function in the export of proteins via homooligomeric structures.

2. Computer programs

Computer programs used were as follows: (1) the PSI-BLAST program [20] with iterations to convergence was used to screen the databases for homologues of the five OMPs that represent the focus of this study. The query sequences were those included on our web site (see www-biology.ucsd.edu/~msaier/transport/). The homologues found and reported in this review represent those proteins in the databases as of June–July, 2001. (2) The CLUSTAL X program [21] and (3) the TREE program [22] were used for multiple alignment of homologous sequences and derivation of phylogenetic trees with the aid of the BLOSUM30 scoring matrix and the TREEVIEW drawing program [23] (see Ref. [24] for evaluation of these and other relevant programs). (4) The TMPred program [25] and (5) the TopPred2 program [26] were used for prediction of the integral membrane topologies of individual proteins. (6) The DAS program was used for prediction of secondary structure. (7) The WHAT program [27], with a sliding window of from 7 to 21 residues, was used to simultaneously predict hydropathy, amphipathicity (angle of 100° for α-helix; angle of 180° for β-strand), topology and secondary structure of individual proteins. (8) The AveHAS program [28] was used for plotting average hydropathy, similarity and amphipathicity as a function of alignment position in the multiple alignments. These programs are available on our "software" and "biotools" web sites http://www-biology.ucsd.edu/~msaier/transport/ and http://www-biology.ucsd.edu/~yzhai/biotools.html, respectively).

In this paper, we use the WHAT [27] and AveHAS [28] programs in combination to predict transmembrane β-strands in porins. These recently developed programs combine several established programs to make structural predictions about transmembrane proteins. For example, the WHAT program examines individual proteins, using JNET [29] and MEMSAT [30] for secondary structure and transmembrane topology prediction, respectively. Both of these programs are among the best available for these purposes. The AveHAS program first generates a multiple alignment for a collection of homologous sequences [21] and then averages (1) hydropathy, (2) amphipathicity and (3) similarity plots to provide structural information that is much more reliable than possible when evaluating single proteins sequences [28]. Transmembrane β-strands can thus be accurately predicted because they exhibit (1) predicted β-structure using JNET, (2) increased hydrophobicity, relative to other portions of the polypeptide chain, and (3) increased amphipathicity when the angle is set at 180° as is appropriate for β-strands [31]. This method predicts transmembrane β-strands with 70–85% accuracy. For example, for the following outer membrane β-barrel proteins of known three-dimensional structure (all from *Escherichia coli*), OmpF (PDB code #1opf) has 16 β-strands but the program predicts 12 (75%). LamB (PDB #1mal) has 18 β-strands; 13 (72%) are predicted; FepA (PDB #1by5) has 22; 19 (86%) are predicted; the OmpX protein (PDB #1qi9) has 8; 6 (75%) are predicted; and phospholipase A (PDB #1qd6) has 12; 10 (83%) are predicted.

3. The fimbrial usher protein (FUP) family (TC #1.B.11)

The FUP family consists of a group of large proteins (most in the 800–900 amino acyl residue (aa) range) present in the outer membranes of Gram-negative bacteria, cyanobacteria and *Deinococcus radiodurans* (Table 1; [13,32,33]). They are believed to contain a large central domain that spans the membrane 24 times as β-strands, presumably forming a β-barrel structure and a transmembrane pore [12,34]. They also possess N-terminal and C-terminal periplasmic domains which may function in protein folding and subunit assembly [14,35]. Each FUP acts in the assembly process together with a periplasmic fimbrial chaperone protein [12,34,36]. The mechanism by which the assembled fimbrial structure is exported through the usher protein across the outer membrane is not well understood.

A single bacterial species such as *E. coli* may be capable of synthesizing numerous fimbriae, and the operon encoding the structural proteins of each fimbrium also encodes the fimbrium-specific periplasmic chaperone protein and the fimbrium-specific outer membrane usher protein [12,37]. Phylogenetic analyses suggest that the chaperone protein and the usher protein, in general, evolved in parallel from their evolutionary precursor proteins [13].

One member of the FUP family, PapC (Table 1), has been shown to form oligomeric channels, 2 nm in diameter, in the outer membrane of *E. coli* [38]. This pore size is large enough to accommodate fimbrial subunits and even partially assembled linear structures. Complexes formed by members of the FUP family may be superficially similar to complexes formed by the PulD secretin (see Table 7) and other related proteins involved in secretion across Gram-negative bacterial outer membranes (C. Peabody, M.-R. Yen, Y.J. Chung and M.H. Saier, Jr., unpublished—in prep.).

Table 1 lists the currently sequenced FUP family members according to organism. With the exception of *D. radiodurans*, an organism classified as Gram-positive [39], all proteins are derived from recognized Gram-negative

Table 1
Recognized homologues of the FUP family

Abbreviation	Database description	Organism	Bacterial type	Size	GI number
FimC Bpe	outer membrane usher protein FimC precursor	*Bordetella pertussis*	β	873	462099
Orf Dra	hypothetical protein	*Deinococcus radiodurans*	Dei	729	10957506
AfaC Eco	outer membrane usher protein AfaC precursor	*Escherichia coli*	γ1	859	1703198
AggC Eco	outer membrane usher protein AggC precursor	*Escherichia coli*	γ1	842	1168385
CfaC Eco	CFA/I fimbrail subunit C precursor	*Escherichia coli*	γ1	869	116127
CooC Eco	CooC protein precursor	*Escherichia coli*	γ1	872	2121085
CotC Eco	CotC protein precursor	*Escherichia coli*	γ1	866	2121089
CS3-2 Eco	outer membrane usher protein CS3-2 precursor	*Escherichia coli*	γ1	937	1169100
CsdD Eco	outer membrane usher protein CssD precursor	*Escherichia coli*	γ1	819	1706159
CshB Eco	CshB porin (usher)	*Escherichia coli*	γ1	800	2808451
CssD Eco	outer membrane usher protein CssD precursor	*Escherichia coli*	γ1	802	1706160
CswD Eco	CS12 fimbria outer membrane usher protein precursor	*Escherichia coli*	γ1	835	13096074
FaeD Eco	outer membrane usher protein FaeD precursor	*Escherichia coli*	γ1	812	119815
FanD Eco	outer membrane usher protein FanD precursor	*Escherichia coli*	γ1	783	119821
FasD Eco	outer membrane usher protein FasD precursor	*Escherichia coli*	γ1	835	1169651
HtrE Eco	outer membrane usher protein HtrE precursor	*Escherichia coli*	γ1	865	2506411
PmfC Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	821	2851668
PapC Eco	outer membrane usher protein PapC precursor	*Escherichia coli*	γ1	836	129618
SfmD Eco	outer membrane usher protein SfmD precursor	*Escherichia coli*	γ1	867	2494481
YbgQ Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	818	2829628
YehB Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	826	465572
YagX Eco	hypothetical 91.2 kDa protein	*Escherichia coli*	γ1	841	2495503
YcbS Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	866	2829634
FocD Eco	outer membrane usher protein FocD precursor	*Escherichia coli*	γ1	875	1169721
FimD Eco	outer membrane usher protein FimD precursor	*Escherichia coli*	γ1	878	729491
Orf1 Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	881	3915996
Orf2 Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	838	1176812
Orf3 Eco	hypothetical outer membrane usher protein	*Escherichia coli*	γ1	793	1176192
Orf4 Eco	putative FUP	*Escherichia coli*	γ1	883	12515165
Orf5 Eco	putative fimbrial chaperone	*Escherichia coli*	γ1	807	1850972
Orf6 Eco	putative fimbrial usher	*Escherichia coli*	γ1	844	12518578
Orf7 Eco	putative fimbrial usher	*Escherichia coli*	γ1	879	12516702
HifC Hin	outer membrane usher protein HifC precursor	*Haemophilus influenzae*	γ2	837	1170260
MrkC Kpn	outer membrane usher protein MrkC precursor	*Klebsiella pneumoniae*	γ1	828	127306
Orf Mlo	hypothetical protein	*Mesorhizobium loti*	α	807	13476410
PhfD Plu	PhfD protein (partial)	*Photorhabdus luminescens*	γ1	799	13236169
PmfC Pmi	outer membrane usher protein PmfC precursor	*Proteus mirabilis*	γ1	828	1709669
AtfC Pmi	outer membrane usher protein	*Proteus mirabilis*	γ1	843	1504107
MrpC Pmi	MrpC protein	*Proteus mirabilis*	γ1	871	485956
Orf1 Pae	hypothetical protein PA4652	*Pseudomonas aeruginosa*	γ3	790	11350238
Orf2 Pae	probable fimbrial biogenesis usher protein PA0994	*Pseudomonas aeruginosa*	γ3	839	11351298
Orf3 Pae	probable fimbrial biogenesis usher protein PA4084	*Pseudomonas aeruginosa*	γ3	895*	11351300
Orf4 Pae	probable fimbrial biogenesis usher protein PA2130	*Pseudomonas aeruginosa*	γ3	872	11351299
TofC Sen	TcfC protein	*Salmonella enterica*	γ1	895*	5640161
SefC Sen	outer membrane usher protein SefC precursor	*Salmonella enteritidis*	γ1	814	464755
LpfC Sty	outer membrane usher protein LpfC precursor	*Salmonella typhimurium*	γ1	842	1170817
FimD Sty	outer membrane usher protein FimD precursor	*Salmonella typhimurium*	γ1	870	585135
BcfC Sty	bovine colonization factor BcfC	*Salmonella typhimurium*	γ1	870	4530570
PefC Sty	outer membrane usher protein PefC precursor	*Salmonella typhimurium* plasmid pCRR10	γ1	802	585660
Orf Ssp	hypothetical protein	*Synechocystis* sp.	Cya	892	7469533
CsuD Vpa	CsuD protein	*Vibrio parahaemolyticus*	γ1	802	13649959
Orf Xfa	outer membrane usher protein precursor XF0081	*Xylella fastidiosa*	γ4	901	11277504
MyfC Yen	outer membrane usher protein MyfC precursor	*Yersinia enterocolitica*	γ1	841	462676
PsaC Ype	outer membrane usher protein PsaC precursor	*Yersinia pestis*	γ1	831*	2506412
Caf1A Ype	F1 capsule anchoring protein precursor	*Yersinia pestis*	γ1	833	115438
Orf1 Ype	hypothetical protein	*Yersinia pestis*	γ1	863	11277505
Orf2 Ype	F1 capsule anchoring protein	*Yersinia pestis* plasmid pMT1	γ1	833	3747030
PsaC Yps	outer membrane usher protein PsaC precursor	*Yersinia pseudotuberculosis*	γ1	832*	2494482

* These proteins are reported on the database to be smaller than reported here, usually due to incorrect initiation codon assignment.

bacteria. *D. radiodurans* exhibits an unusual dual membrane envelope where the two membranes are of essentially the same lipid composition. The outer membrane of this organism lacks lipopolysaccharide, the cell surface antigenic hallmark of Gram-negative bacteria [39].

Examination of Table 1 reveals an unexpected organismal representation. Thus, of the 58 proteins listed, more than half (30) are from *E. coli* strains, and 16 more are from the closely related enteric γ-proteobacterial genuses, *Klebsiella* (one protein), *Proteus* (three proteins), *Salmonella* (six proteins) and *Yersinia* (six proteins). Thus, only 12 proteins are from more divergent bacteria. Of these, four are from *Pseudomonas aeruginosa*, and one each is derived from *Haemophilus influenzae*, *Photorhabdus luminescens*, *Vibrio parahaemolyticus* and *Xylella fastidiosa*, all non-enteric γ-proteobacteria. The remaining four proteins are from *Bordetella pertussis*, a β-proteobacterium, *Mesorhizobium loti*, an α-proteobacterium, *Synechocystis* sp., a cyanobacterium, and *D. radiodurans*. Many bacterial kingdoms that include organisms with completely sequenced genomes, including (1) the Spirochetes, (2) Neisserial species, (3) *Chlamydia*, (4) *Helicobacter*, (5) *Rickettsia*, (6) *Mycoplasma*, (7) low G + C Gram-positive bacteria, (8) high G + C Gram-positive bacteria and (9) primitive bacteria such as *Thermatoga* and *Aquifex*, do not exhibit a FUP family homologue. Thus, FUP family members appear to be largely restricted to the proteobacteria, and very few homologues are found outside of this bacterial kingdom.

The protein size variation recorded in Table 1 is noteworthy. The smallest protein (729 aas) is derived from *D. radiodurans*, while the largest (937 aas) is the CS3-2 protein of *E. coli*, an unusual FUP family homologue with a C-terminal tail that exhibits no sequence similarity to anything else in the current databases. Excluding these two proteins, the size range varies from 783 aas to 901 aas, and the two largest of these proteins are from *X. fastidiosa* (901 aas) and *Synechocystis* sp. (892 aas). The *Salmonella enterica* TofC protein (889 aas) is the next largest homologue.

A multiple alignment of all identified FUP family homologues revealed only a single fully conserved residue, a glycine. However, only conservative substitutions were observed at many positions. The most highly conserved region occurred at alignment positions 405–426 in a central amphipathic β-sheet region. From this region, we sought to derive both a consensus sequence (the majority residue(s) at any position are portrayed) and a signature sequence. The consensus sequence for this region was:

> Q N G (Y R) $(\text{L I V})_2$ Y X_3 (L I V) $(\text{P A S})_2$ G* (P A) F X (L I V) X D (L I V)
> (Residues in parentheses represent alternative possibilities at a single position;
> X = any residue; G* = the fully conserved glycine).

Our attempts to derive a FUP family-specific signature sequence were not successful.

Average hydropathy, amphipathicity and similarity plots were generated using the AveHAS program [28]. The hydropathy plot revealed that excluding the N-terminal hydrophobic leader sequence, no region exhibited a sufficiently long stretch of strongly hydrophobic residues to pass through the membrane as an α-helix. Using an angle of 180° (as is appropriate for β-strands) for the derivation of the average amphipathicity plot, and a window size of seven residues, many short peaks of average amphipathicity in the second two-thirds of these proteins corresponded to peaks of average hydrophobicity as well as peaks of similarity. Many of these peaks in the central domain may correspond to transmembrane β-strands that contribute to the integral membrane β-barrel structure. Such structures are characteristic of outer membrane porins [40].

The WHAT program [27] was used with representative FUP family members to estimate topology and secondary structure. The program predicted that FUP family proteins are predominantly of β-structure throughout their lengths. It was therefore not possible to easily predict where the boundaries between the transmembrane domains and the putative periplasmic, C-terminal, β-structured, hydrophilic domains occur.

The FUP family phylogenetic tree, based on the CLUSTAL X-generated multiple alignment, is shown in Fig. 1A. As noted above, all but a few of the members of the FUP family are derived from γ-proteobacteria closely related to the enteric bacteria. Most of the proteins can be considered to fall into 10 clusters. γ-Proteobacterial proteins are found in all 10 clusters, and all remaining branches bear a single protein outside of the γ-proteobacterial clusters. These include the proteins from *D. radiodurans* (Dra), the α-proteobacterium *M. loti* (Mlo) and the cyanobacterium, *Synechocystis* sp. (Ssp). The β-proteobacterial (*Bordetella*) protein (Bpe) is found as an outlier of cluster 10. Thus, all four proteins from non-γ-proteobacteria are on branches distant from the other proteins. The α-proteobacterial protein (Mlo) is possibly orthologous to the cluster 9 proteins while the β-proteobacterial protein (Bpe) is possibly orthologous to cluster 10 proteins.

Fig. 1B shows the phylogenetic tree for the 16S ribosomal RNAs of the organisms that include FUP family homologues. Seven organisms (*E. coli*, *Salmonella typhimurium*, *Klebsiella pneumoniae*, *P. luminescens*, *Proteus mirabilis*, *Yersinia pestis* and *V. parahaemolyticus*) form a tight cluster at the top left of the tree. *H. influenzae* is the next closest relative, while *P. aeruginosa* is significantly more divergent. The three more distant proteobacteria represented in Fig. 1B are more closely related to the aforementioned γ-proteobacteria than they are to the two non-proteobacteria, as expected. In this connection, it is interesting to note that all but one of the seven organisms outside of the tight γ1 cluster exhibit only one protein of the FUP family per organism, the exception being *P. aeruginosa*, which has an exceptionally large genome. Fully sequenced genomes are available for many

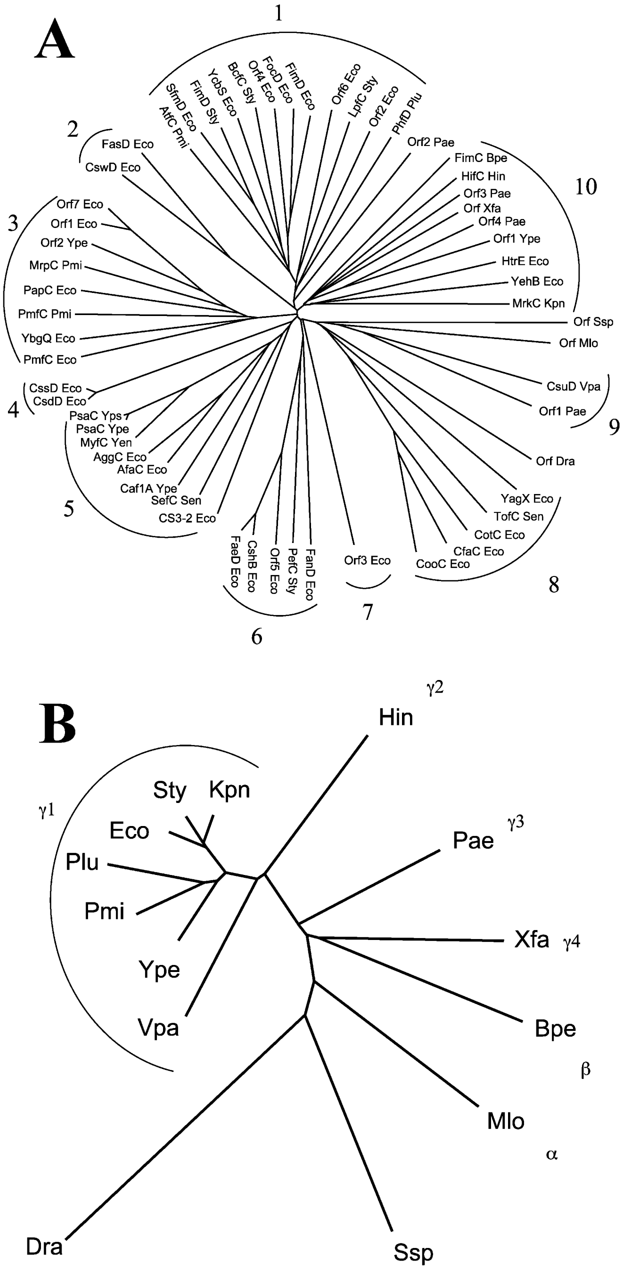

Fig. 1. The fimbrial usher protein (FUP) family. (A) Phylogenetic tree for currently recognized members of the FUP family. (B) 16S rRNA tree for organisms known to possess one or more FUP family homologue(s). Protein abbreviations are as indicated in Table 1. Greek letters in (B) and in subsequent figures refer to the proteobacterial subgroup.

of these organisms. It is interesting that both *Buchneria* sp. and *Vibrio cholera*, two γ-proteobacteria for which fully sequenced genomes were available at the time these studies were conducted, lack a FUP family member. It should be noted that many *E. coli* paralogues are plasmid-encoded and/or specific to particular strains of this species.

The FUP family proteins were divided into three approximately equal sized fragments, the N-terminal, central and C-terminal thirds, and these were analyzed phylogenetically. The resultant trees were strikingly similar to each other and to the tree shown in Fig. 1A, with just one minor exception (data not shown). The CS3-2 Eco protein thirds 1 and 2 clustered as shown in Fig. 1A, in cluster 5, but third 3 clustered loosely with PhfD Plu and Orf2 Eco in cluster 1. Because the CS3-2 Eco fragment sequences were always found to branch from points near the centers of these unrooted trees, it was not possible to establish that the latter difference was statistically significant. Thus, our phylogenetic analyses did not reveal obvious shuffling of protein domains during the evolutionary divergence of FUP family members.

4. The outer membrane factor (OMF) family (TC #1.B.17)

Proteins of the OMF family [15,17] function in conjunction with a primary cytoplasmic membrane transporter of the MFS (TC #2.A.1) [41], the ABC superfamily (TC #3.A.1) [42], the RND superfamily (TC #2.A.6) [43] and the PET family (TC #9.B.4) [44], as well as a membrane fusion protein (MFP; TC #8.A.1) [45]. The complex thus formed allows transport (export) of various solutes (heavy metal cations; drugs, oligosaccharides, proteins, etc.) across the two membranes of the Gram-negative bacterial cell envelope in a single energy-coupled step. The OMF proteins probably form homotrimeric, 12-stranded, β-barrel-type pores in the outer membrane through which the solutes pumped out of the cytoplasm or cytoplasmic membrane pass in response to the energy-coupled export process catalyzed by the cytoplasmic membrane permease [46]. One of these proteins, TolC of *E. coli* (Table 2), has been purified as a trimer, crystallized in two-dimensional lattices by reconstitution in phospholipid bilayers, and shown at 12 Å resolution to exhibit three-fold symmetry with an outer diameter of 58 Å and an internal stain-filled pore [46]. In one case, the complex of primary transporter, MFP and OMF was shown to form transiently in response to substrate binding [47].

The crystal structure of *E. coli* TolC has more recently been solved to 2.1 Å resolution [48]. Three TolC protomers form a continuous, solvent-accessible conduit, a channel tunnel over 140 Å long, which spans both the outer membrane (as 12 β-strands, four each per protomer) and the periplasmic space (as 12 α-helices, six continuous, six discontinuous, four each protomer). The α-helices are continuous with the β-strands. In the crystal structure, the periplasmic end of the tunnel is sealed by sets of coiled helices that might untwist upon contact with the primary permease to open the channel.

The OMFs exhibit a pseudosymmetrical structure due to the presence of two internally duplicated segments, and the outer membrane β-barrel is assembled from the three protomers with each one contributing four β-strands. Each β-strand is between 10 and 13 residues long. The strands both curve and twist, yielding a superhelical structure, but the channel is wide open and fully accessible to solvent. The possibility of channel closure due to conformational mobility has not been excluded [48]. The results clearly suggest that the OMFs (and not the MFPs) are largely responsible for the formation of both the trans-outer membrane and trans-periplasmic channels [49,50]. The functional roles played by the MFPs have yet to be determined.

Table 2 lists the currently recognized members of the OMF family; 102 proteins are tabulated. Of these proteins, all are derived from Gram-negative bacteria with the sole exception of the dual membrane-possessing Gram-positive bacterium, *D. radiodurans*. Thirty-two bacterial genuses are represented, and these organisms include α-, β-, γ- and ε-proteobacteria, spirochetes, the cyanobacterium *Synechocystis* sp., *D. radiodurans*, *Porphyromonas gingivalis*, and *Aquifex aeolicus*. Several species exhibit multiple paralogues. For example, *P. aeruginosa* has 18, *E. coli* and *Caulobacter crescentus* have seven each, *A. aeolicus* has six and *V. cholera* has five. Nevertheless, it is worth noting that a few Gram-negative bacteria with fully sequenced genomes (species of *Buchneria* and *Chlamydia*, *Thermatoga maritima* and *Treponema pallidum*) lack a recognizable OMF family member.

A multiple alignment of all of the OMF family homologues was generated with the CLUSTAL X program. Although many positions were well conserved, none was fully conserved, and at no single position did conservative substitutions occur exclusively. A signature sequence could not be derived for this family. Like FUP family members, the proteins of the OMF family exhibit fairly uniform similarity throughout much of their lengths, although some proteins exhibited internal insertions relative to their homologues. For example, Orf Ssp exhibits a segment of about 60 residues (alignment positions 45–105) following the hydrophobic leader sequence (alignment positions 15–40), which is not found in any other homologue. Alignment positions 110–155 proved to be well conserved in all homologues, but Orf1 Hpy and Orf12 Pae exhibit extensions at alignment positions 155–215 that are not present in the other homologues. Alignment positions 235–356 and 383–430, as well as positions 460–585 are also well conserved among almost all homologues. However, the smallest OMF family member, from *D. radiodurans*, exhibits two internal deletions (alignment positions 150–250 and 421–473), relative to all other homologues, both in regions of poor conservation where alignment gaps are common in many homologues as noted by Koronakis et al. [48]. It is therefore clear that OMFs exhibit extended regions of strong conservation as well as interdomain linker regions where length variability is common.

Because the high-resolution three-dimensional structure of the *E. coli* TolC is known [48], these observations can be placed into a structural framework. TolC consists of an internally repeated structure. Each repeat unit consists of two short helices (H1 and H2 or H5 and H6), followed by two β-strands (S1 and S2 or S4 and S5), followed by two long helices (H3 and H4 or H5 and H6), followed by a short β-strand (S3 or S6). Within the repeat units, only S3 and S6 are not demonstrably homologous to each other. The major gaps in the aligned sequences occur (1) between the leader sequence and H1, (2) in the extracellular loop between S1 and S2, (3) in the junctional region between the two halves of the proteins (between S3 and H5), (4) in the second extracellular loop between S4 and S5, and (5) following H6. As noted above, S3 and S6 are not homologous. Thus, the extended regions encompassing H1 + H2 + S1 and the homologous H5 + H6 + S4 are well conserved without insertions or deletions. The same is true of the extended regions encompassing S2 + H3 + H4 and the homologous S5 + H5 +H6. These results show that each repeat unit consists of two indivisible units, each containing two α-helices and one β-strand where insertions and deletions cannot (or do not) occur. The only exception is the homologue in *D. radiodurans*.

The phylogenetic tree for the OMF family is shown in Fig. 2A, and that for the 16S ribosomal RNA tree of the represented organisms is shown in Fig. 2B. The latter tree shows seven separate clusters of proteobacteria plus five divergent bacterial species [51].

The tree for the OMFs reveals 12 clusters plus many proteins that do not cluster significantly with any other protein. Several observations are worthy of note: (1) γ-proteobacterial proteins are found in all clusters except clusters 3 and 11, and they are also found on many divergent branches not included in the 12 clusters. Thus, many close and many distant paralogues are found in organisms that display multiple OMFs. (2) Clusters 3 and 11 include only α-proteobacterial proteins. Moreover, proteins from these bacteria are also found in clusters 6, 8 and 12, as well as on distant branches outside of the 12 clusters. However, none of these proteins clusters closely with a protein from another group of proteobacteria, leading to the conclusion that horizontal transfer of genes encoding OMFs to α-proteobacteria from bacteria of other groups has not occurred in recent evolutionary history. (3) β-Proteobacterial proteins are found in clusters 6, 7, 9 and 12, and in all such cases, loose clustering with a *Pseudomonas* protein is observed. Proteins from β-proteobacteria are also found on non-clustering branches. As for the α-proteobacterial proteins, evidence for horizontal transfer of the encoding genes is lacking. (4) The single *X. fastidiosa* protein is found in cluster 2, loosely associated with both β- and γ-proteobacterial proteins, while the two *Stenotrophomonas maltophilia* proteins localize to cluster 8, loosely associated with one protein from *E. coli* and another from *C. crescentus*. (5) ε-Proteobacterial proteins are found in clusters 2 and 5, but they are distant members

Table 2
Recognized homologues of the OMF family

Abbreviation	Database description	Organism	Bacterial type	Size	GI number
NodT Atu	NodT homolog	*Agrobacterium tumefaciens*	α	484	9957271
NccC Ade	nickel–cobalt–cadmium resistance protein NccC precursor	*Alcaligenes denitrificans*	β	437	3914124
CzcC Asp	divalent cation resistant determinant protein C	*Alcaligenes* sp.	β	417	2120972
Orf1 Aae	conserved hypothetical protein aq_1332	*Aquifex aeolicus*	Aqu	415	7514442
Orf2 Aae	hypothetical protein aq_1059	*Aquifex aeolicus*	Aqu	417	7517364
Orf3 Aae	conserved hypothetical protein aq_699	*Aquifex aeolicus*	Aqu	437	7514526
Orf4 Aae	hypothetical protein aq_1093	*Aquifex aeolicus*	Aqu	425	7517373
Orf5 Aae	hypothetical protein aq_1670	*Aquifex aeolicus*	Aqu	402	7517470
Orf6 Aae	hypothetical protein aq_1133	*Aquifex aeolicus*	Aqu	392	7517382
CyaE Bpe	CyaE protein precursor	*Bordetella pertussis*	β	474	117799
Orf Bbu	hypothetical protein BB0142	*Borrelia burgdorferi*	Spi	440	7463239
OpcM Bce	OpcM	*Burkholderia cepacia*	β	512	1061410
FusA Bce	fusaric acid resistance protein FusA precursor	*Burkholderia cepacia*	β	530	9911073
Orf Bps	unknown	*Burkholderia pseudomallei*	β	541	4139248
SapF Cfe	SapF	*Campylobacter fetus*	ε	433	4009449
Orf1 Cje	probable outer membrane channel protein Cj0365c	*Campylobacter jejuni*	ε	492	11347034
Orf2 Cje	probable outer membrane protein Cj0608	*Campylobacter jejuni*	ε	456	11347036
Orf3 Cje	probable outer membrane component of efflux system Cj1031	*Campylobacter jejuni*	ε	424	11347035
TolC Ccr	outer membrane protein TolC, putative	*Caulobacter crescentus*	α	483	13422661
RsaF Ccr	type I secretion system outer membrane protein RsaF	*Caulobacter crescentus*	α	527	13422305
Orf1 Ccr	efflux system protein	*Caulobacter crescentus*	α	467	13422053
Orf2 Ccr	efflux system protein	*Caulobacter crescentus*	α	478	13423215
Orf3 Ccr	efflux system protein	*Caulobacter crescentus*	α	500	13424871
Orf4 Ccr	metal ion efflux outer membrane factor protein family	*Caulobacter crescentus*	α	412	13424310
Orf5 Ccr	metal ion efflux outer membrane factor protein family	*Caulobacter crescentus*	α	421	13423923
Orf Dra	hypothetical protein	*Deinococcus radiodurans*	Dei	347	7472106
TolC Eae	TolC protein	*Enterobacter aerogenes*	γ1	486	13539234
PrtF Eam	PrtF protein	*Erwinia amylovora*	γ1	437	4826418
PrtF Ech	protease secretion protein PrtF precursor	*Erwinia chrysanthemi*	γ1	462	131076
TolC Eco	outer membrane protein TolC precursor	*Escherichia coli*	γ1	495	135980
IbeB Eco	IbeB protein	*Escherichia coli*	γ1	460	4835717
Orf1 Eco	putative outer membrane channel protein	*Escherichia coli*	γ1	457	13361330
CusC Eco	probable outer membrane lipoprotein CusC precursor	*Escherichia coli*	γ1	457	2495560
Orf2 Eco	putative outer membrane export protein	*Escherichia coli*	γ1	451	12513363
YjcP Eco	hypothetical outer-membrane lipoprotein YjcP precursor	*Escherichia coli*	γ1	488	2851560
YohG Eco	hypothetical outer-membrane lipoprotein YohG precursor	*Escherichia coli*	γ1	478	9911117
Orf1 Hin	hypothetical protein HI1462	*Haemophilus influenzae*	γ2	454	1175810
Orf2 Hin	hypothetical protein HI1340	*Haemophilus influenzae*	γ2	441	1175736
Orf1 Hpy	hypothetical protein jhp1382	*Helicobacter pylori*	ε	510	7465023
Orf2 Hpy	hypothetical protein jhp0552	*Helicobacter pylori*	ε	477	7464752
Orf3 Hpy	hypothetical protein jhp0905	*Helicobacter pylori*	ε	431	7464875
Hel Lpn	Hel protein	*Legionella pneumophila*	γ3	414	511474
NodT Mlo	outer membrane protein, NodT candidate	*Mesorhizobium loti*	α	466	13471200
TolC Nme	secretion protein, probable NMB1737	*Neisseria meningitidis*	β	467	11354143
Orf Nme	multidrug efflux pump channel protein NMB1714	*Neisseria meningitidis*	β	467	11353796
Orf Pmu	unknown	*Pasteurella multocida*	γ2	455	12720790
IbeB Pmu	IbeB	*Pasteurella multocida*	γ2	463	12722419
PG41 Pgi	immunoreactive 52 kDa antigen PG41	*Porphyromonas gingivalis*	Por	462	5759281
PG53 Pgi	immunoreactive 50 kDa antigen PG53	*Porphyromonas gingivalis*	Por	444	5759289
PG52 Pgi	immunoreactive 51 kDa antigen PG52	*Porphyromonas gingivalis*	Por	455	5759287
ZapD Pmi	ZapD	*Proteus mirabilis*	γ1	449	3493599
AprF Pae	alkaline protease secretion protein AprF PA1248	*Pseudomonas aeruginosa*	γ3	481	416635

Table 2 (*continued*)

Abbreviation	Database description	Organism	Bacterial type	Size	GI number
Orf1 Pae	probable secretion protein PA4974	*Pseudomonas aeruginosa*	γ3	482	11351822
OprM Pae	outer membrane protein OprM precursor PA0427	*Pseudomonas aeruginosa*	γ3	485	12644685
Orf2 Pae	probable secretion protein PA3404	*Pseudomonas aeruginosa*	γ3	451	11351819
Orf3 Pae	probable secretion protein PA4144	*Pseudomonas aeruginosa*	γ3	471	11351821
Orf4 Pae	probable outer membrane lipoprotein precursor PA2525	*Pseudomonas aeruginosa*	γ3	498	11351561
OprJ Pae	outer membrane protein OprJ precursor PA4597	*Pseudomonas aeruginosa*	γ3	479	12230972
Orf5 Pae	probable outer membrane protein PA2837	*Pseudomonas aeruginosa*	γ3	479	11351567
Orf6 Pae	probable outer membrane protein PA2391	*Pseudomonas aeruginosa*	γ3	474	11351565
OprN Pae	outer membrane protein OprN precursor PA2495	*Pseudomonas aeruginosa*	γ3	472	11350716
Orf7 Pae	probable outer membrane efflux protein precursor PA4208	*Pseudomonas aeruginosa*	γ3	487	11351558
Orf8 Pae	probable outer membrane protein PA5158	*Pseudomonas aeruginosa*	γ3	492	11351574
Orf9 Pae	hypothetical protein PA3894	*Pseudomonas aeruginosa*	γ3	496	11350029
CzcC Pae	outer membrane protein precursor CzcC PA2522	*Pseudomonas aeruginosa*	γ3	428	11350718
Orf10 Pae	hypothetical protein PA1875	*Pseudomonas aeruginosa*	γ3	425	11349339
Orf11 Pae	probable outer membrane efflux protein precursor PA3521	*Pseudomonas aeruginosa*	γ3	491	11351557
Orf12 Pae	hypothetical protein PA4592	*Pseudomonas aeruginosa*	γ3	493	11350212
Orf13 Pae	probable outer membrane component of multidrug efflux pump PA1238	*Pseudomonas aeruginosa*	γ3	482	11351556
AprF Pbr	AprF protein	*Pseudomonas brassicacearum*	γ3	453	9438191
OMP Pch	outer membrane protein	*Pseudomonas chlororaphis*	γ3	453	6013393
TliF Pfl	ABC transporter TliF	*Pseudomonas fluorescens*	γ3	481	4063019
CztC Pfl	CztC protein	*Pseudomonas fluorescens*	γ3	406	12484564
AprF Pfl	zinc-protease transporter	*Pseudomonas fluorescens*	γ3	471	2952089
TtgC Ppu	outer membrane channel protein	*Pseudomonas putida*	γ3	484	8163737
AggA Ppu	agglutination protein	*Pseudomonas putida*	γ3	452	281563
SrpC Ppu	outer membrane channel protein	*Pseudomonas putida*	γ3	470	2605915
TtgF Ppu	outer membrane channel protein	*Pseudomonas putida*	γ3	480	6912016
OprM Psy	putative outer membrane efflux protein OprM	*Pseudomonas syringae*	γ3	478	10764639
EprF Pto	EprF protein	*Pseudomonas tolaasii*	γ3	488	3646415
CzcC Rme	cobalt–zinc–cadmium resistance protein CzcC precursor	*Ralstonia metallidurans*	β	417	2507004
CnrC Rme	nickel and cobalt resistance protein CnrC precursor	*Ralstonia metallidurans*	β	418	729165
NodT1 Rle	nodulation protein T precursor	*Rhizobium leguminosarum*	α	467	128491
NodT2 Rle	hypothetical protein 471	*Rhizobium leguminosarum*	α	471	541015
NodT3 Rle	nodulation protein T precursor	*Rhizobium leguminosarum*	α	482	462727
TolC Rpr	outer membrane protein TolC precursor RP224	*Rickettsia prowazekii*	α	456	7467903
TolC Sen	outer membrane protein TolC precursor	*Salmonella enteritidis*	γ1	491	2495191
SilC Sty	probable outer membrane lipoprotein SilC precursor	*Salmonella typhimurium*	γ1	461	13633958
Orf Sty	ABC exporter outer membrane component homolog	*Salmonella typhimurium*	γ1	439	7467234
HasF Smar	HasA export system outer membrane protein HasF	*Serratia marcescens*	γ1	500	11277508
LipD Smar	LipD protein	*Serratia marcescens*	γ1	464	3080540
Orf Sar	putative aromatic efflux pump outer membrane protein	*Sphingomonas aromaticivorans*	α	483	10956846
SmeC Smal	SmeC protein	*Stenotrophomonas maltophilia*	γ4	471	5764626
SmeF Smal	outer membrane protein	*Stenotrophomonas maltophilia*	γ4	466	11071585
Orf Ssp	hypothetical protein slr1270	*Synechocystis* sp.	Cya	526	7470402
TolC EBA	predicted outer membrane protein TolC	Uncultured proteobacterium EBAC31A08	γ3	442	9971916
TolC1 Vch	outer membrane protein TolC precursor VC2436	*Vibrio cholerae*	γ1	438	11135318
TolC2 Vch	probable outer membrane protein TolC VC1565	*Vibrio cholerae*	γ1	419	11355953
Orf1 Vch	agglutination protein VC1621	*Vibrio cholerae*	γ1	445	11354392
Orf2 Vch	conserved hypothetical protein VC1606	*Vibrio cholerae*	γ1	476	11354620
Orf3 Vch	probable multidrug resistance protein VC1409	*Vibrio cholerae*	γ1	484	11355941
Orf Xfa	outer membrane export factor XF2586	*Xylella fastidiosa*	γ4	452	11277506

of these clusters. The majority of proteins from these organisms do not cluster with any other protein. (6) Finally, all proteins from the five bacteria that do not cluster on the 16S rRNA tree do not cluster in the OMF tree. These observations, taken together, suggest that there has been little or no late horizontal transfer of genes encoding OMFs from the α-, β-, γ- or ε-proteobacteria to any other group of these organisms or to the more

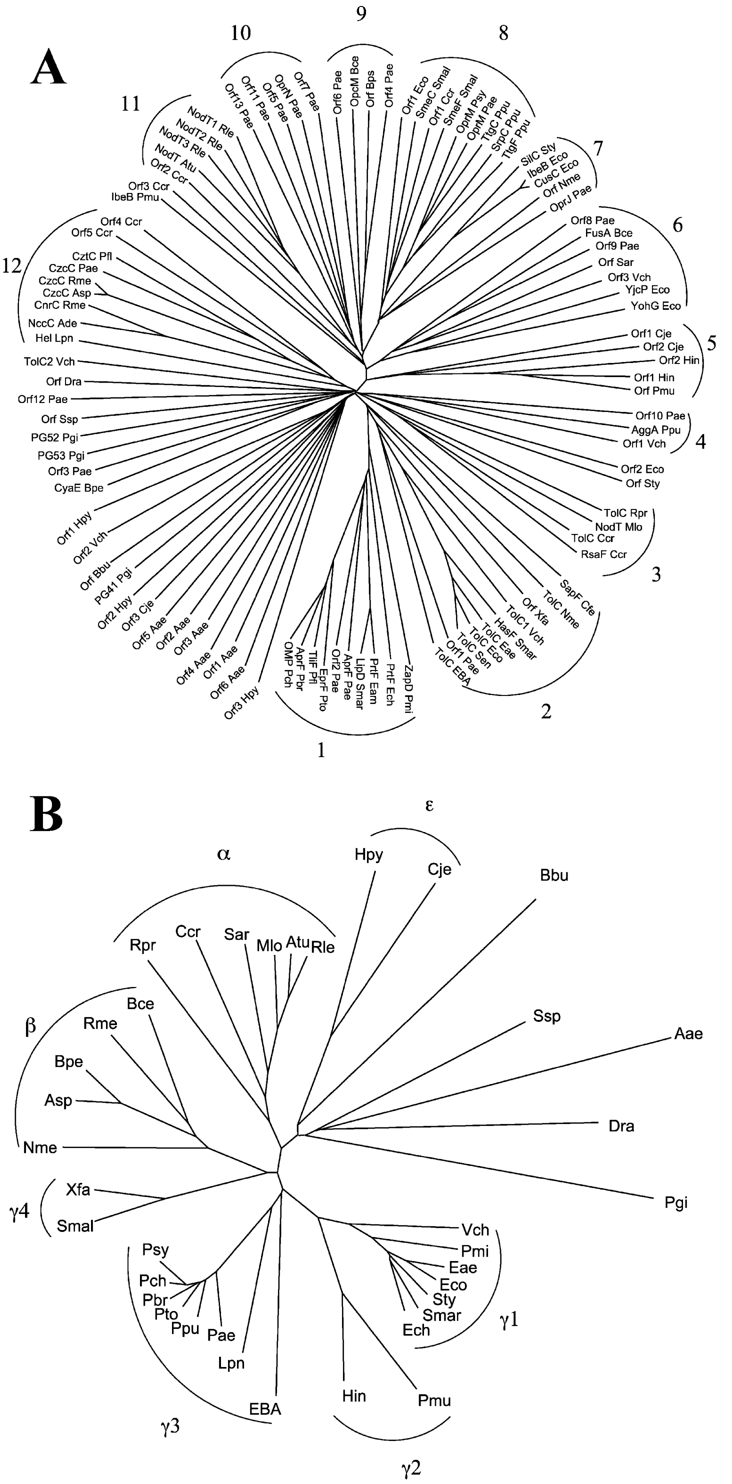

Fig. 2. The outer membrane factor (OMF) family. (A) Phylogenetic tree for currently recognized members of the OMF family. (B) 16S rRNA tree for organisms known to possess one or more OMF family homologue(s). Protein abbreviations are as indicated in Table 2.

divergent bacterial types in which these homologues are found. We suggest that unlike the FUP family, the OMF family may be an ancient one that dates back before divergence of the major bacterial kingdoms. (7) Although they do not cluster with other proteins, two of the three paralogues from *P. gingivalis* cluster loosely together. Moreover, all of the six *A. aeolicus* paralogues are found adjacent to each other on the OMF tree, and all three *Helicobacter pylori* paralogues as well as the three *Campylobacter jejuni* paralogues are only distantly related to

Table 3
Recognized homologues of the AT family

Abbreviation	Database description	Organism	Bacterial type	Size	GI number	S	β
EstA Asp	esterase	*Acidiphilium* sp. AIU409	α	627	4704345	+	+
Pert Bbr	pertactin precursor	*Bordetella bronchiseptica*	β	911	400749	+	+
Pert Bpa	pertactin precursor	*Bordetella parapertussis*	β	922	129828	+	+
BapB Bpe	BapB Protein	*Bordetella pertussis*	β	482	10944730	–	+
BrkA Bpe	BrkA protein	*Bordetella pertussis*	β	1010	2120986	+	+
Vag8 Bpe	Vag8 protein	*Bordetella pertussis*	β	915	29997419	+	+
TcfA Bpe	tracheal colonization factor A precursor	*Bordetella pertussis*	β	672	2121002	+	+
Pert Bpe	pertactin precursor	*Bordetella pertussis*	β	910	464364	+	+
BapC Bpe	putative autotransporter	*Bordetella pertussis*	β	759	3411270	+	+
Phg Bpe	Phg protein	*Bordetella pertussis*	β	418	8670938	+	+
BapA Bpe	BapA protein	*Bordetella pertussis*	β	903	10944728	+	+
PmpD Cmu	polymorphic membrane protein D family TC0197	*Chlamydia muridarum*	Chla	1520	11362550	+	+
PmpD Ctr	probable outer membrane protein D	*Chlamydia trachomatis*	Chla	1531	7468993	+	+
Orf Cab	putative 98 kDa outer membrane protein	*Chlamydophila abortus*	Chla	926	1657778	+	+
Pmp Cpn	polymorphic membrane protein D family CP0897	*Chlamydophila pneumoniae*	Chla	1609	7468524	+	+
Pmp10 Cpn	probable outer membrane protein Pmp10 precursor CP0303	*Chlamydophila pneumoniae*	Chla	928	14195016	+	+
Pmp8 Cpn	probable outer membrane protein Pmp8 precursor	*Chlamydophila pneumoniae*	Chla	930	14195066	+	+
Pmp18 Cpn	polymorphic outer membrane protein e/f family	*Chlamydophila pneumoniae*	Chla	946	7468498	+	+
EspC Eco	enterotoxin EspC	*Escherichia coli*	γ1	1305	11527908	+	+
Orf1 Eco	putative beta-barrel outer membrane protein	*Escherichia coli*	γ1	1349	12513130	+	+
YdeK Eco[a]	YdeK protein	*Escherichia coli*	γ1	1325[b]	1787788	+	–
Orf2 Eco	hypothetical protein b1509	*Escherichia coli*	γ1	466	7466188	–	+
Orf6 Eco[a]	putative ATP-binding component of a transport system	*Escherichia coli*	γ1	556[b]	1787416	+	–
Orf3 Eco	hypothetical protein b1170	*Escherichia coli*	γ1	347[b]	7466147	–	+
YfaL Eco	hypothetical 131.2 kDa protein	*Escherichia coli*	γ1	1250	2506696	+	+
Sat Eco	secreted autotransporter toxin	*Escherichia coli*	γ1	1295	11096073	+	+
EspP Eco	serine protease EspP	*Escherichia coli*	γ1	1300	10955344	+	+
YejA Eco	hypothetical 98.4 kDa protein	*Escherichia coli*	γ1	1569	2507221	+	+
Tsh Eco	Tsh protein	*Escherichia coli*	γ1	1377	2126101	+	+
YejO Eco	hypothetical 91.2 kDa protein	*Escherichia coli*	γ1	863	465619	+	+
Orf4 Eco	putative flagellin structural protein	*Escherichia coli*	γ1	980	13359881	+	+
YaiT Eco[a]	YaiT protein precursor	*Escherichia coli*	γ1	486	1786569	+	–
YaiU Eco	hypothetical 50.3 kDa protein	*Escherichia coli*	γ1	467	2495526	–	+
TibA Eco	TibA protein	*Escherichia coli*	γ1	989	5305639	+	+
Orf5 Eco	probable membrane protein b1202	*Escherichia coli*	γ1	955	7466752	+	+
AG43 Eco	antigen 43 precursor	*Escherichia coli*	γ1	1039	2506898	+	+
AidA-I Eco	adhesin AIDA-I precursor	*Escherichia coli* plasmid pIB6	γ1	1286	543788	+	+
YchA Eco	adhesin AidA-I precursor	*Escherichia coli* plasmid F	γ1	1399[b]	9507741	+	+
YcbB Eco	adhesin AidA-I precursor	*Escherichia coli* plasmid F	γ1	1769[b]	9507739	+	+
EaaA Eco	EaaA protein	*Escherichia coli* prophage P-EibA	γ1	1335	7532795	+	+
IgA Hin	immunoglobulin A1 protease precursor	*Haemophilus influenzae*	γ2	1694	1170513	+	+
Hap Hin	adhesion and penetration protein precursor	*Haemophilus influenzae*	γ2	1394	1170167	+	+
Hap1 Hin	adhesion and penetration protein precursor	*Haemophilus influenzae*	γ2	1409	1170166	+	+
Hsr Hmu	major ring-forming surface protein precursor	*Helicobacter mustelae*	ε	1519	1086005	+	+
VacA1 Hpy	vacuolating cytotoxin precursor	*Helicobacter pylori*	ε	1290	2499106	+	+
VacA2 Hpy	vacuolating cytotoxin precursor	*Helicobacter pylori*	ε	1288	12230793	+	+
Orf1 Hpy	toxin-like outer membrane protein HP0922	*Helicobacter pylori*	ε	2529	7465392	+	+
Orf2 Hpy	toxin-like outer membrane protein HP0289	*Helicobacter pylori*	ε	2893	7465390	+	+
Orf4 Hpy[a]	hypothetical protein HP0609	*Helicobacter pylori*	ε	1238	7464312	+	–
Orf3 Hpy	toxin-like outer membrane protein HP0610	*Helicobacter pylori*	ε	1943	7465391	–	+
Orf Kas	hypothetical protein; (fragment)	*Kluyvera ascorbata*	γ1	652	9843779	+	+
Ssa1 Mha	serotype-specific antigen 1 precursor	*Mannheimia haemolytica*	γ2	932	401120	+	+
Orf1 Mlo	serine proteinase	*Mesorhizobium loti*	α	1213	13471834	+	+
Orf2 Mlo	hypothetical protein	*Mesorhizobium loti*	α	1008	13471533	+	+
Orf3 Mlo	hypothetical protein	*Mesorhizobium loti*	α	3659	13471072	+	+

(continued on next page)

Table 3 (*continued*)

Abbreviation	Database description	Organism	Bacterial type	Size	GI number	S	β
IgA Ngo	IgA-specific serine endopeptidase precursor	*Neisseria gonorrhoeae*	β	1532	124244	+	+
Orf1 Nme	probable virulence associated protein NMA1725	*Neisseria meningitidis*	β	656[b]	11354121	+	+
VapA Nme	probable virulence associated protein NMA2175	*Neisseria meningitidis*	β	679[b]	11354122	+	+
IgA Nme	IgA-specific metalloendopeptidase NMB0700	*Neisseria meningitidis*	β	1815	11353752	+	+
App Nme	adhesion and penetration protein NMB1985	*Neisseria meningitidis*	β	1457	11280386	+	+
Orf2 Nme	serine-type peptidase NMB1998	*Neisseria meningitidis*	β	1431	11354147	+	+
Orf3 Nme	Serotype-1-specific antigen, probable NMB1969	*Neisseria meningitidis*	β	1082	11354148	+	+
Orf1 Pmu	unknown	*Pasteurella multocida*	γ2	850	12722129	+	+
Orf2 Pmu	unknown	*Pasteurella multocida*	γ2	1080	12721328	+	+
NanB Pmu	sialidase NanB	*Pasteurella multocida*	γ2	1070	11464736	+	+
Est Pmu	Est protein	*Pasteurella multocida*	γ2	679	12720285	+	+
Lip1 Plu	lipase1 precursor	*Photorhabdus luminescens*	γ3	645	729942	+	+
Orf1 Pae	probable serine proteinase PA3535	*Pseudomonas aeruginosa*	γ3	995	11351832	+	+
EstA Pae	esterase EstA PA5112	*Pseudomonas aeruginosa*	γ3	646	11348487	+	+
Orf2 Pae	hypothetical protein PA0328	*Pseudomonas aeruginosa*	γ3	647	11348838	+	+
PspB Pbr	PspB homolog	*Pseudomonas brassicacearum*	γ3	1030	9438192	+	+
PspB Pfl	serine protease homologue	*Pseudomonas fluorescens*	γ3	1036	4115629	+	+
PrtB Pfl	PrtB protien	*Pseudomonas fluorescens*	γ3	1036	8895500	+	+
PspA Pfl	serine protease homologue	*Pseudomonas fluorescens*	γ3	985	4115628	+	+
Ytrp Ppu	hypothetical 62.7 kDa protein (fragment)	*Pseudomonas putida*	γ3	592	732298	+	+
EprS Pto	serine protease	*Pseudomonas tolaasii*	γ3	985	3646417	+	+
OmpB Rae	OmpB; fragment	*Rickettsia aeschlimannii*	α	1617	6969926	–	+
OmpB Raf	OmpB; fragment	*Rickettsia africae*	α	1616	6969928	–	+
OmpB Rak	OmpB; fragment	*Rickettsia akari*	α	1619	6969930	–	+
OmpB Rau	OmpB; fragment	*Rickettsia australis*	α	1620	6969934	–	+
OmpA Rau	outer membrane protein A	*Rickettsia australis*	α	2106	11641393	+	+
OmpB Rco	OmpB; fragment	*Rickettsia conorii*	α	1617	6969958	–	+
OmpB Rhe	OmpB; fragment	*Rickettsia helvetica*	α	1604	6969966	–	+
OmpB Rho	OmpB; fragment	*Rickettsia honei*	α	1616	6969964	–	+
OmpB Rja	outer membrane protein B precursor	*Rickettsia japonica*	α	1656	6685710	+	+
OmpB Rma	OmpB; fragment	*Rickettsia massiliae*	α	1616	6969944	–	+
OmpB1 Rmo	OmpB; fragment	*Rickettsia mongolotimonae*	α	1616	6969946	–	+
OmpB2 Rmo	OmpB; fragment	*Rickettsia montanensis*	α	1615	6969948	–	+
OmpB Rpa	OmpB; fragment	*Rickettsia parkeri*	α	1616	6969950	–	+
OmpB Rpr	outer membrane protein B precursor	*Rickettsia prowazekii*	α	1643	6685725	+	+
Sca3 Rpr	Cell surface antigen (sca3) RP451	*Rickettsia prowazekii*	α	2340	7467598	+	+
OmpB Rrh	OmpB; fragment	*Rickettsia rhipicephali*	α	1616	6969954	–	+
OmpB Rri	outer membrane protein B precursor	*Rickettsia rickettsii*	α	1654	6685726	+	+
190K Rri	190 kDa ANTIGEN precursor	*Rickettsia rickettsii*	α	2249	112710	+	+
OmpB1 Rsp	OmpB; fragment	*Rickettsia* sp. A-167	α	1614	6969932	–	+
OmpA Rsp	OmpA; fragment	*Rickettsia* sp. HLJ-054	α	1058	9789172	–	+
OmpB2 Rsp	OmpB; fragment	*Rickettsia* sp. S	α	1615	6969956	–	+
OmpB Rty	outer membrane protein B precursor	*Rickettsia typhi*	α	1645	3023209	+	+
SapA Sty	SapA protien	*Salmonella typhi*	γ1	961	10945146	+	+
MisL Sty	MisL protien	*Salmonella typhimurium*	γ1	955	4324610	+	+
ShdA Sty	ShdA protien	*Salmonella typhimurium*	γ1	2035	5107805	+	+
ApeE Sty	outer membrane esterase	*Salmonella typhimurium*	γ1	656	2896133	+	+
BigA Sty	putative surface-exposed virulence protein BigA	*Salmonella typhimurium*	γ1	1963	5081595	+	+
PrtS Sma	extracellular serine protease precursor	*Serratia marcescens*	γ1	1045	131087	+	+
PrtT Sma	extracellular serine protease precursor	*Serratia marcescens*	γ1	1045	266848	+	+
SSP-h1 Sma	SSP-h1	*Serratia marcescens*	γ1	1036	3688585	+	+
SSP-h2 Sma	serine proteinase h2	*Serratia marcescens*	γ1	1034	7435686	+	+
SepA Sfl	secreted protease	*Shigella flexneri*	γ1	1364	13449013	+	+
Pic Sfl	Pic protien	*Shigella flexneri* 2a	γ1	1373	12643212	+	+
SigA Sfl	exported serine protease SigA	*Shigella flexneri* 2a	γ1	1285	7682555	+	+
Sap Sfl	Sap protein	*Shigella flexneri* 2a	γ1	1040	12643222	+	+

Table 3 (*continued*)

Abbreviation	Database description	Organism	Bacterial type	Size	GI number	S	β
VirG Sfl	VirG protein	*Shigella flexneri* plasmid pMYSH6000	γ1	1102	96922	+	+
Orf1 Xfa	serine proteinase XF1851	*Xylella fastidiosa*	γ4	1000	11362667	+	+
Orf2 Xfa	serine proteinase XF0267	*Xylella fastidiosa*	γ4	999[b]	11362665	+	+
Orf3 Xfa	serine proteinase XF1026	*Xylella fastidiosa*	γ4	1002[b]	11362666	+	+
Orf4 Xfa	lipase/esterase XF0781	*Xylella fastidiosa*	γ4	597	11362229	+	+
YapA Ype	YapA protein	*Yersinia pestis*	γ1	1432	10945150	−	+
YapB Ype	YapB protein (partial)	*Yersinia pestis*	γ1	1052	10945152	+	+
YapC Ype	YapC protein	*Yersinia pestis*	γ1	638	10945154	+	+
YapD Ype	YapD protein	*Yersinia pestis*	γ1	1457	10945156	−	+
YapE Ype	YapE protein	*Yersinia pestis*	γ1	1072	10945158	+	+
YapF Ype	YapF protein	*Yersinia pestis*	γ1	761	10945160	+	+
YapG Ype	YapG protein	*Yersinia pestis*	γ1	994	10945162	+	+
YapH Ype	YapH protein	*Yersinia pestis*	γ1	3705	10945164	+	+

[a] The two proteins within a bracket (() are believed to correspond to a passenger protein (upper protein) for the autotransporter domain localized to the C-terminus of the protein below it (lower protein).

[b] These proteins differ in length from those presented in the database due to the identification of additional regions. The database entries are believed to have resulted from incorrect initiation codon assignment or to sequencing errors.

the other proteins although loose clustering of ε-proteobacterial proteins is often observed. It therefore appears that these paralogues arose by early gene duplication events in their respective bacterial lineages, arguing against horizontal transfer of genes encoding OMFs.

5. The autotransporter (AT) family (TC #1.B.12)

Pathogenic Gram-negative bacteria produce a diversity of virulence factors which cross the cytoplasmic membrane via the Sec (general secretory) pathway (TC #3.A.5), and following cleavage of their N-terminal targeting sequence, they enter the periplasm of the Gram-negative bacterial cell envelope [52–55]. The C-terminal 250–300 amino acyl residues of proteins known as "autotransporters" fold and insert into the outer membrane to give rise to putative β-barrel structures with 14 transmembrane β-strands (TMSs) [18,56–58]. This structure presumably forms a pore through which the N-terminal virulence factor is transported to the extracellular milieu [59]. Pore formation in lipid bilayers by one of these AT domains, that in the BrkA protein of *B. pertussis* (Table 3; TC #1.B.12.2.3), has been demonstrated [60]. Following its export, the precursor virulence factor is usually (but not always) proteolytically digested to release a soluble protein that can promote virulence [61].

Although the C-terminal AT domains are all homologous, they are extremely diverse in sequence [18]. Moreover, the N-terminal virulence factor domains are not all homologous. These various protein domains can (1) catalyze proteolysis, (2) serve as adhesins, (3) mediate actin-promoted bacterial motility or (4) serve as cytotoxins to animal cells (Tables 3 and 4). The intact protein, prior to processing, can vary in size between 681 and 1546 amino acyl residues. A lack of specificity for the protein transported has been demonstrated for some AT [62].

Table 3 presents the 120 members of the AT family identified in the current databases. They are derived from 20 bacterial genuses. These genuses include members of the α-, β-, γ- and ε-proteobacteria, but only one bacterial kingdom outside of the proteobacteria, the chlamydial kingdom, is represented. Several organisms possess multiple paralogues including *E. coli* (22), *Y. pestis* (8), *B. pertussis* (8), *Neisseria meningitidis* (6) and *H. pylori* (5).

The size variation observed for ATs is tremendous: the smallest homologue is Orf3 Eco (347 aas) and consists only of the AT domain. An upstream gene in the same operon (Orf6 Eco) encodes the putative passenger or toxin protein. Two *E. coli* homologues have 466–487 aas, and both of these (Orf2 Eco and YaiU Eco) have the putative passenger proteins (YdcK and YaiT) encoded by distinct genes mapping directly upstream of the AT domain-containing protein. *H. pylori* contains a large AT (Orf3 Hpy) in an operon with an upstream gene encoding a probable passenger protein (Orf4 Hpy). Thus, it appears that the passenger protein is not always covalently linked to the AT. It is possible, however, that one or more of these examples, where the passenger protein and the AT domain exist as two distinct polypeptide chains, are abnormal. For example, *E. coli* YaiT and YaiU are separated by the insertion sequence IS3. The largest AT homologues are YapH Ype (3705 aas) and Orf3 Mlo (3659 aas). In these and all other cases, the AT domains are at the extreme C-termini of the proteins.

The 120 AT domains were multiply aligned. No residue was fully conserved, but two residues were particularly well conserved. The G at alignment position 154 was conserved in all but seven proteins and the P at position 306 was conserved in all but three proteins. Only one of the exceptional proteins lacking a G at position 154 also lacked a P at position 306. Two other residues, G at position 349 and G at position 424 were also well (but less well) conserved. Thus, the best-conserved residues are all structural residues.

Average hydropathy, amphipathicity and similarity plots were generated from the complete AT domain multiple alignment. In agreement with previous results [18] where only 18 proteins were examined, 14 peaks of hydrophobicity proved to exhibit amphipathic character typical of transmembrane β-strands, and all proved to be well conserved (see our web site).

The phylogenetic tree for the AT family is shown in Fig. 3A while the 16S rRNA tree for the represented organisms is shown in Fig. 3B. The latter tree reveals that a preponder-

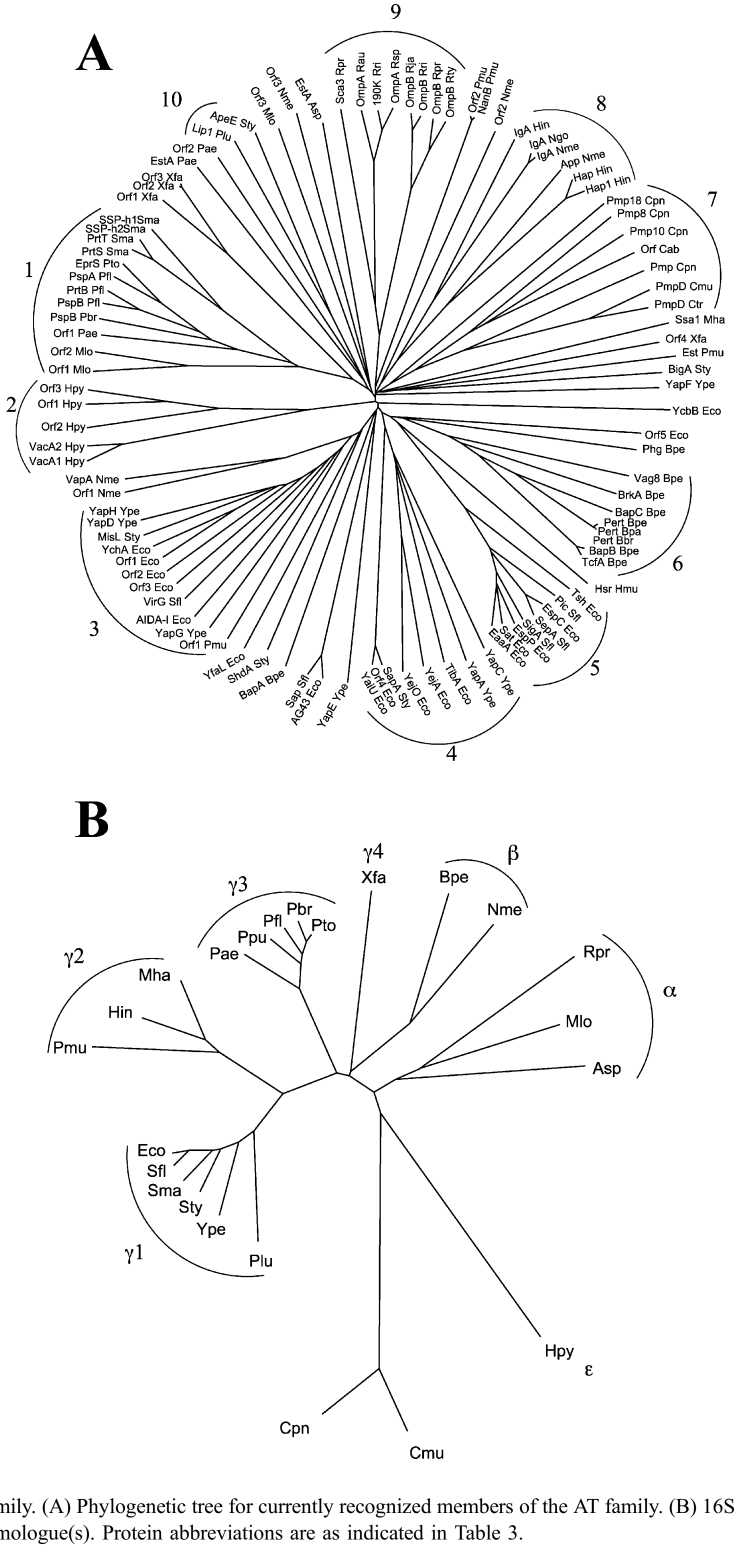

Fig. 3. The Autotransporter (AT) family. (A) Phylogenetic tree for currently recognized members of the AT family. (B) 16S rRNA tree for organisms known to possess one or more AT family homologue(s). Protein abbreviations are as indicated in Table 3.

ance of organisms known to possess AT domains are included in the four clusters of γ-proteobacteria (labeled γ1, γ2, γ3 and γ4, respectively), as well as the α, β and ε subdivisions of the proteobacteria. As noted above, only one group of non-proteobacterial organisms, the chlamydial group, exhibits AT domains.

The AT domain phylogenetic tree, shown in Fig. 3A, reveals 33 deep-rooted branches, but only 10 of these branches display protein clustering. In almost every case, each such branch includes proteins from organisms that belong to a single phylogenetic group. Thus, proteins from γ1-proteobacteria are exclusively present on 10 branches as well as in three clusters that include proteins from other proteobacterial groups. γ2-Proteobacterial proteins are found on six branches as well as four deep-rooted branches bearing only one protein. Two other branches include γ1- and β-proteobacterial proteins.

γ3-Proteobacterial proteins are found exclusively on three branches while γ4-proteobacterial proteins are found exclusively in two clusters. α-Proteobacterial proteins are found on three branches, and two of these include proteins only from these organisms. β-Proteobacterial proteins are found on seven branches, only one of which also has proteins from another group (γ2). The ε-proteobacterial proteins are exclusively localized to two branches while the proteins from the chlamydial group are exclusively found on one deep-rooted branch.

Summarizing these observations, 29 of the 33 branches bear proteins from a single organismal type, with only four bearing proteins from two or more types (Table 4). Moreover, proteins in most clusters exhibit a uniform size and function (Table 4) showing that phylogeny provides a reliable guide to function. It seems clear that close homologues arose almost exclusively by speciation and late gene duplication events within a single organism; horizontal transfer of genes encoding ATs between distant organismal types was a rare evolutionary event.

Table 4
Size variation correlated with organismal source and putative function for the 10 clusters of autotransporters (see Fig. 3A)

Cluster[a]	Organisms represented[b]	Average size ± S.D.[c]	(Putative) function of substrate protein[d]
1	γ1, γ3, α	1036 ± 63	protease
2	ε	2146 ± 793	cytotoxin
3	γ1, γ2	1435 ± 804	adhesin
4	γ1	1048 ± 303	uncertain (enterotoxin)[e]
5	γ1	1329 ± 38	protease
6	β	822 ± 174	adhesin
7	*Chlamydia*	1198 ± 333	cytolysin
8	γ2, β	1150 ± 170	adhesin/protease
9	α	1899 ± 318	surface antigen
10	γ1, γ3	650 ± 8	hydrolase (lipase, esterase, peptidase)

[a] Cluster refers to the phylogenetic cluster shown in Fig. 3A.

[b] All Greek letter entries refer to the category of proteobacteria from which the proteins are derived.

[c] Sizes are expressed in numbers of amino acyl residues ± S.D.

[d] The known or postulated functions of the N-terminal substrate protein domains that are believed to be exported via the C-terminal β-barrel-forming autotransporter domains are provided. Outside of the 10 major clusters can be found several of the functional types described for the major clusters as well as other functional types such as putative ATPases, flagellins and sialidases (see Table 3).

[e] N-terminal substrate protein domains in cluster 4 are sequence divergent and are annotated in the databases as (a) adhesins, (b) proteases, (c) ATP-binding proteins and (d) flagellins. However, based on our analyses, some of these assignments are likely to be incorrect.

6. The two-partner secretion (TPS) family (TC #1.B.20)

The first member of the TPS family to be characterized was the ShlB (HlyB) protein of *Serratia marcescens*, which exports the ShlA hemolysin from the periplasm of the Gram-negative bacterial envelope into the external medium [63]. ShlA reaches the periplasm by export from the cytoplasm via the GSP or IISP (TC #3.A.5). ShlB and some, but not all, TPS homologues include domains with both an outer membrane export channel and a "hemolysin activator." ShlB activates ShlA by derivatization with phosphatidyl ethanolamine [64].

Several ShlB homologues have been functionally characterized [65–68]. The channel activities of some of these homologues have been demonstrated [69,70], and topological features of these putative β-barrel porins have been studied. One such protein, FhaC of *B. pertussis*, exhibits a surface-exposed N-terminus and an odd number of β-strands with large surface loops and small periplasmic loops [70,71].

Substrates of bacterial TPS family exporters include Ca^{2+}-independent cytolysins, an iron acquisition protein and several adhesins. Specificity with respect to particular protein substrates has been demonstrated [66]. The hallmarks of TPS systems are the presence of (1) an N-proximal module where specific secretion signals in the substrate protein are found and (2) a β-barrel channel (TpsB) homologue [19]. Usually, the genes encoding these two proteins occur within an operon. After transport of the unfolded protein across the cytoplasmic membrane via the GSP, the substrate protein probably folds in the periplasm and/or on the periplasmic surface of the outer membrane before it is exported via the TPS porin [19].

Sequenced protein members of the TPS family retrieved from the current databases are listed in Table 5. Of the five outer membrane protein-translocating porin types characterized in this report, TPS family members are the most widespread in nature even though there are fewer sequenced members than in the OMF, AT and secretin families. Thus, only 77 TPS homologues were identified, but in addition to all of the major subgroups of proteobacteria except the δ subgroup, they were found in chlamydia, cyanobacteria, *D. radiodurans*, *A. aeolicus*, *Fusobacterium necrophorum* and *T. maritima*. Although homologues were not identified in archaea, they were represented in the animal, plant and fungal kingdoms of eukaryotes.

Table 5
Recognized homologues of the TPS family

Abbreviation	Database description	Organism	Bacterial type	Size	GI number
Prokaryotic					
Omp Aae	outer membrane protein	*Aquifex aeolicus*	Aqu	778	7520765
Orf Aae	hypothetical protein aq_050	*Aquifex aeolicus*	Aqu	861*	7517352
FhaC Bbr	FhaC protein	*Bordetella bronchiseptica*	β	583	6650632
FhaC Bpe	hemolysin activator-like protein FhaC precursor	*Bordetella pertussis*	β	584	462082
Omp1 Bme	OMP1 precursor	*Brucella melitensis*	α	782	1262291
Orf1 Cje	outer membrane protein Cj0129c	*Campylobacter jejuni*	ε	739	11346784
Orf2 Cje	probable outer-membrane protein Cj0975	*Campylobacter jejuni*	ε	574	11347039
Orf1 Ccr	outer membrane protein	*Caulobacter crescentus*	α	769	13423368
Orf2 Ccr	conserved hypothetical protein	*Caulobacter crescentus*	α	628	13423000
Orf3 Ccr	hypothetical protein	*Caulobacter crescentus*	α	513	13421807
Orf Cmu	outer membrane protein, probable TC0512	*Chlamydia muridarum*	Chla	792	11362439
Omp85 Ctr	probable omp85 analog	*Chlamydia trachomatis*	Chla	792	7468991
Omp85 Cpn	Omp85 analog	*Chlamydophila pneumoniae*	Chla	790	7468478
Orf Dra	outer membrane protein	*Deinococcus radiodurans*	Dei	846	7473239
EthB Eta	activation/secretion protein EthB	*Edwardsiella tarda*	γ1	559	11360479
Orf1 Eco	hypothetical protein b0177	*Escherichia coli*	γ1	810	2506737
Orf2 Eco	putative outer membrane transporter	*Escherichia coli*	γ1	539	12514411
Orf3 Eco	hypothetical 64.8 kDa protein	*Escherichia coli*	γ1	577	732290
Orf Fne	unknown; fragment	*Fusobacterium necrophorum*	Fus	338	13469803
LspB Hdu	hemolysin accessory protein homolog (Fragment)	*Haemophilus ducreyi*	γ2	474	7467544
HhdB Hdu	hemolytic protein hhdB precursor	*Haemophilus ducreyi*	γ2	532	7467546
D15 Hin	protective surface antigen D15 precursor	*Haemophilus influenzae*	γ2	797	1169202
HxuB2 Hin	heme/hemopexin utilization protein B precursor	*Haemophilus influenzae*	γ2	565	1170439
HxuB1 Hin	heme/hemopexin utilization protein B precursor	*Haemophilus influenzae*	γ2	565	1170438
Orf1 Hin	putative accessory processing protein	*Haemophilus influenzae*	γ2	545	475772
Orf2 Hin	hypothetical protein HI0698 precursor	*Haemophilus influenzae*	γ2	578	1176923
D15 Hpy	protective surface antigen D15	*Helicobacter pylori*	ε	916	7465335
Orf1 Mlo	outer membrane protein	*Mesorhizobium loti*	α	794	13470835
Orf2 Mlo	hypothetical protein	*Mesorhizobium loti*	α	626*	13471632
Omp85 Ngo	outer membrane protein	*Neisseria gonorrhoeae*	β	792	1766042
HecB1 Nme	hemolysin activation protein HecB, NMB1762	*Neisseria meningitidis*	β	595	11353175
HecB2 Nme	hemolysin activation protein HecB, NMB1780	*Neisseria meningitidis*	β	580	11353176
Orf1 Nme	hemolysin activator-related protein NMB0496	*Neisseria meningitidis*	β	559	7413434
Omp85 Nme	outer membrane protein Omp85 NMB0182	*Neisseria meningitidis*	β	797	11279714
Orf2 Nme	conserved hypothetical protein NMB2134	*Neisseria meningitidis*	β	635	11282853
LspB1 Pmu	LspB protein	*Pasteurella multocida*	γ2	576*	12720262
LspB2 Pmu	LspB protein	*Pasteurella multocida*	γ2	573*	12720265
Orf1 Pmu	putative hemolysin activator-like protein; fragment	*Pasteurella multocida*	γ2	482	7716521
Orf2 Pmu	unknown	*Pasteurella multocida*	γ2	791	12722432
Oma87 Pmu	outer membrane antigen Oma87	*Pasteurella multocida*	γ2	789	1401350
Orf3 Pmu	unknown	*Pasteurella multocida*	γ2	586	12722231
HecB Pch	HecB protein	*Pectobacterium chrysanthemi*	γ1	558	1772622
Orf Plu	outer membrane antigen	*Photorhabdus luminescens*	γ1	797	5689866
HpmB Pmi	hemolysin activator protein precursor	*Proteus mirabilis*	γ1	561	123203
Orf1 Pae	conserved hypothetical protein PA0040	*Pseudomonas aeruginosa*	γ3	562	11347607
Orf2 Pae	hypothetical protein PA2463	*Pseudomonas aeruginosa*	γ3	565	11349581
Orf3 Pae	probable outer membrane protein PA3648	*Pseudomonas aeruginosa*	γ3	797	11351570
Orf4 Pae	hypothetical protein PA4540	*Pseudomonas aeruginosa*	γ3	545	11350202
Orf5 Pae	hypothetical protein PA4624	*Pseudomonas aeruginosa*	γ3	568	11350221
Orf6 Pae	hypothetical protein PA0692	*Pseudomonas aeruginosa*	γ3	544	11348954
Orf7 Pae	conserved hypothetical protein PA2543	*Pseudomonas aeruginosa*	γ3	579	11347901
Orf8 Pae	hypothetical protein PA3339	*Pseudomonas aeruginosa*	γ3	728	11349874
Omp1 Rpr	outer membrane protein Omp1 RP160	*Rickettsia prowazekii*	α	768	7467902
HlyB Smar	hemolysin activator protein precursor	*Serratia marcescens*	γ1	557	123205
Orf Ssp	hypothetical protein slr1661	*Synechocystis* sp.	Cyan	654	7470479
Iap75 Ssp	chloroplast import-associated channel IAP75	*Synechocystis* sp.	Cyan	861	7469855
Orf Tma	hypothetical protein	*Thermotoga maritima*	The	711	7462447
Orf1 Vch	surface antigen VC2252	*Vibrio cholerae*	γ1	803	11279712
Orf2 Vch	conserved hypothetical protein VC2548	*Vibrio cholerae*	γ1	582	11282638
Orf3 Vch	hypothetical protein VC1749	*Vibrio cholerae*	γ1	408	11346255

Table 5 (*continued*)

Abbreviation	Database description	Organism	Bacterial type	Size	GI number
Prokaryotic					
Orf Xor	putative outer membrane protein	*Xanthomonas oryzae*	γ4	593	11693113
Orf1 Xfa	outer membrane hemolysin activator protein XF2550	*Xylella fastidiosa*	γ4	597	11362429
Orf2 Xfa	outer membrane antigen XF1046	*Xylella fastidiosa*	γ4	784	11279711
Orf3 Xfa	conserved hypothetical protein XF1231	*Xylella fastidiosa*	γ4	617	11360753
Eukaryotic					
Orf1 Ath	gene_id: MOP10.6 — unknown protein	*Arabidopsis thaliana*	Pl	524	10178129
Orf2 Ath	unknown protein	*Arabidopsis thaliana*	Pl	520	6016688
Orf3 Ath	unknown protein	*Arabidopsis thaliana*	Pl	732	13430586
Orf4 Ath	hypothetical protein F26G5.110	*Arabidopsis thaliana*	Pl	435	11357663
Orf5 Ath	outer envelope membrane protein homolog T6H20.230	*Arabidopsis thaliana*	Pl	818	7487986
Orf Cel	hypothetical 43.2 kDa protein C34E10.1	*Caenorhabditis elegans*	An	398	1176527
Orf Dme	hypothetical protein CG7639	*Drosophila melanogaster*	An	463	12585512
CGI51 Has	protein CGI-51	*Homo sapiens*	An	469	12643329
Oep75 Psa	outer envelope membrane protein OEP75 precursor	*Pisum sativum*	Pl	809	1363492
Orf Sce	Ynl026wp	*Saccharomyces cerevisiae*	Fu	484	6324302
Orf Spo	hypothetical 51.8 kDa protein C17C9.06	*Schizosaccharomyces pombe*	Fu	475	1723565

* These proteins are reported to differ in length from that presented due to incorrect initiation codon assignment or sequencing errors.

Examination of Table 5 reveals an interesting distribution of paralogues. Among bacteria: *P. aeruginosa* (8) >*Pasteurrella multocida* (6) >*H. influenzae* (5) >*N. meningitidis* (4) >*E. coli*, *V. cholera*, *X. fastidiosa* and *C. crescentus* (3). *Synechocystis* sp., *A. aeolicus* and *C. jejuni* each have two, while chlamydial species, *H. pylori*, *Rickettsia prowazekii* and *T. maritima* only have one. Several divergent bacterial species with fully sequenced genomes lack homologues. These include all Gram-positive bacteria and spirochetes. Among eukaryotes, only *Arabidopsis thaliana* has more than one paralogue, but it has five. *Saccharomyces cerevisiae*, *S. pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Homo sapiens* each have one.

The sizes of the precursor proteins are tabulated in Table 5. A large fraction of these homologues are in the 500–800 amino acyl residue range, but many are larger, the largest (from *Zymomonas mobilis*) having 1056 residues and the second largest (from *H. pylori*) having 916 residues. While only five proteins have 500–600 aas, 21 have 600–700, seven have 700–800 and only two proteins are larger. The animal and fungal proteins, and one plant protein are among the smallest homologues found (398–484 aas), although most of the plant proteins are of sizes comparable to those found in bacteria (520–818 aas). The size ranges and organismal sources of the six primary clusters (Fig. 4) are presented in Table 6.

The multiple alignment revealed at least 20 clear regions of probable amphipathic β-strands. This observation is in agreement with the documented suggestion that these proteins exist as pore-forming β-barrels. No residue was fully conserved in all of the homologues. However, several features were noteworthy: (1) the N-terminal regions were strongly divergent in sequence, and were consequently excluded from the phylogenetic analyses reported below. (2) Following about 16 putative β-strand regions and preceding the last four such putative strands was a region of high conservation with the following consensus sequence:

(D E) X Hy X Hy G G X_2 (S T) Hy R G (Y F)

(X = any residue; Hy = any hydrophobic residue)

The R G (Y F) motif was conserved in all but nine of the homologues, and in each of these nine proteins, at least two of these three residues were conserved.

The phylogenetic tree for the TPS family is shown in Fig. 4A, and the 16S rRNA tree for represented organisms is shown in Fig. 4B. The latter exhibits organismal clustering as expected with the proteobacteria clustering according to subtype, the six-sequence divergent non-proteobacteria branching from points near the center of the prokaryotic part of the tree, and the eukaryotic part of the tree dividing into three groups: plants, animals and fungi, as expected.

The TPS family tree (Fig. 4A) reveals 20 deep-rooted branches, six of these bearing multiple proteins. One cluster (cluster 1) includes one of the two cyanobacterial homologues plus the outer chloroplast envelope protein, Oep75 of *P. sartorium*, a component of the chloroplast envelope protein import translocase (CEPT) family (TC #3.A.9) [72–76]. Its channel activity has been demonstrated in an artificial lipid bilayer membrane [77]. It is interesting that three *Arabidopsis* paralogues are found in this cluster, but that only one of the two *Synechocystis* paralogues is found therein.

All remaining eukaryotic proteins are found in cluster 2. Except for the two-cluster 2 *A. thaliana* paralogues, eukaryotes with fully sequenced genomes each exhibit a single representative protein in this cluster. Surprisingly, the *C. elegans* protein does not cluster with the other animal proteins. Based on this criterion, it is apparently not an orthologue of the *Drosophila* and human proteins.

Cluster 3 is a large cluster of proteobacterial proteins where no organism has more than one representative pro-

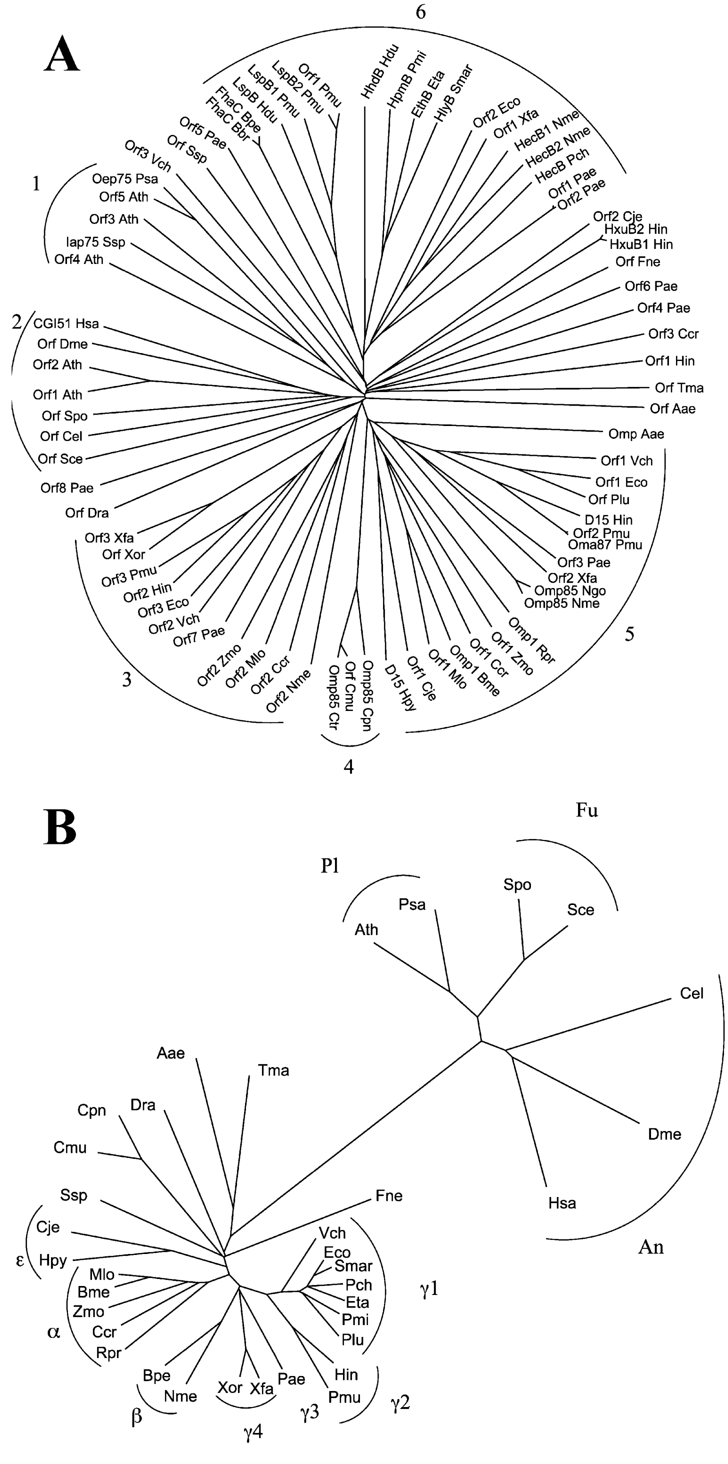

Fig. 4. The two-partner secretion (TPS) family. (A) Phylogenetic tree for currently recognized members of the TPS family. (B) 16S rRNA tree for organisms known to possess one or more TPS family homologue(s). Protein abbreviations are as indicated in Table 5. Pl, plants; Fu, fungi; An, animals.

tein, and every proteobacterial subgroup (γ1, γ2, γ3, γ4, α and β) except δ and ε is represented (Table 6). All of the proteins in cluster 3 exhibit phylogenetic clustering as do the corresponding 16S rRNAs (compare Fig. 4A and B). Further, they are all in the same size range (577–635) except for Orf2 Zmo, which is much larger due to an N-terminal extension of about 200 residues (Tables 5 and 6). Therefore, they are all probably orthologues serving the same function.

Cluster 4 consists of three chlamydial proteins, and no chlamydial protein lies outside of this cluster. They are undoubtedly orthologues.

Table 6
Size variation for the six clusters within the TPS family (see Fig. 4A)

Cluster[a]	Organisms represented[b]	Average size ± S.D.[c]	(Putative) function of substrate protein
1	Eukaryotes and cyanobacteria	805 ± 54	chloroplast proteins
2	Eukaryotes	476 ± 42	?
3	γ1, γ2, γ3, γ4, α, β	612 ± 47	?
4	Chlamydia	791 ± 1	?
5	γ1, γ2, γ3, γ4, α, β, ε	808 ± 70	Zn^{2+}-metaloproteases
6	γ1, γ2, γ3, γ4, β	568 ± 18	cytolysins/adhesins

[a] Cluster refers to the phylogenetic cluster shown in Fig. 4A.

[b] All Greek letter entries refer to the category of proteobacteria from which their proteins are derived.

[c] Sizes are expressed in numbers of amino acid residues ± S.D.

Cluster 5 is a second large cluster of exclusively proteobacterial proteins. The subgroups represented are γ1, γ2, γ3, γ4, β, α, and ε (Table 6). The two very closely related proteins from *P. multocida* are from two different strains. No organism has more than one representative protein in this cluster, and clustering is according to organismal type (i.e., 16S rRNA). Finally, all but three of these proteins are in the same size range (769–810). The exceptions are D15 Hpy, which has an extra internal loop, Orf1 Cje, which has a deletion in a loop, and Orf1 Zmo, which has a 300 residue N-terminal hydrophilic extension. We postulate that like cluster 3, cluster 5 consists exclusively of orthologues serving a common function.

Cluster 6 is a third large cluster of exclusively proteobacterial proteins. The subgroups represented are γ1, γ2, γ3, γ4 and β. In contrast to clusters 3 and 5, clustering is not according to organismal phylogeny. Nevertheless, all of these proteins are in the same size range (532–597) except for LspB Hdu (474) and Orf1 Pmu (482) (see Table 6). Further, within cluster 6, there are three close paralogues from *P. multocida*, two close paralogues from *N. meningitidis* and two close paralogues from *P. aeruginosa*.

Examining paralogues, we find that except for Orfs 1 and 2, all *P. aeruginosa* paralogues are very divergent in sequence. Again, excluding HxuB1 and HxuB2 of *H. influenzae*, all paralogues in this organism are very divergent in sequence. Moreover, the two paralogues in *H. ducreyi* do not have counterparts in *H. influenzae*. *N. meningitidis* has two close paralogues (HecB1 and HecB2) in cluster 6 with its other two paralogues in clusters 3 and 5. Similarly, *P. aeruginosa* has two very close paralogues in cluster 6 but one each in clusters 3 and 5. Finally, the three *E. coli* paralogues and the three *X. fastidiosa* paralogues are in clusters 3, 5 and 6. It appears that while there has been evolutionary pressure to duplicate cluster 6 genes, there has been no pressure to duplicate cluster 3 and 5 genes. Since some proteobacteria with fully sequenced genomes are not represented in these clusters, there must have been a tendency for some of these organisms to lose one or another of these paralogues during evolution.

7. The outer membrane secretin (Secretin) family (TC #1.B.22)

The Secretin family consists of a group of Gram-negative bacterial outer membrane proteins that form multimeric pores through which macromolecules, usually proteins, can pass [78–80]. These proteins form homomultimeric ring structures, 10–20 subunits per complex, with large central pores (inner diameters of 50–100 Å). One secretin, PilQ of *N. meningitidis*, is a dodecamer with 12 identical subunits arranged in a ring [81]. Secretins are large proteins (420–750 amino acyl residues) consisting of two domains: an N-terminal periplasmic domain (the first 280 residues of *Pseudomonas* XcpQ proteins) and a C-terminal "homology" domain that is embedded in the outer membrane. The C-terminal "homology" domains of secretins are exclusively responsible for channel formation [82].

Secretins function in type II protein secretion (TC #3.A.5) (e.g., PulD of *K. oxytoca*), type III protein secretion (TC #3.A.6) (e.g., the hypersensitivity response secretin (HrpH) of *P. syringiae*), host cell invasion (e.g., the protein secretin InvG of *S. typhimurium*), competence (e.g., competence protein E (ComE) of *H. influenzae*), fimbrial protein export and assembly (e.g., the fimbrial assembly protein (PilQ) of *P. aeruginosa*) and phage assembly (e.g., the gene IV protein of bacteriophage f1) [10,83]. In *V. cholerae*, the secretin of the type III secretion system, EpsD, which exports cholera toxin, also exports the filamentous phage, CTXQ, the genome of which encodes cholera toxin [84,85]. Filamentous phage are simultaneously secreted and assembled with coat proteins. The enteropathogenic *E. coli* secretin, BfpB, exports pilin subunits and several PEC proteins, and renders cells sensitive to the antibiotic, vancomycin [86].

Table 7 presents the currently sequenced members of the Secretin family. Ninety-six proteins were identified. These proteins are primarily from Gram-negative bacteria although several are from plasmids and phage of *E. coli* and *P. aeruginosa*, and one is from the purported Gram-positive *D. radiodurans*, which has two membranes of similar composition. Organisms with large numbers of secretin paralogues include *P. aeruginosa* with eight paralogues, *E. coli* and *M. loti*, both with six paralogues, *V. cholerae* with five and *Pectobacterium chrysanthemi* with four. Many additional organisms have two or three (Table 7). There is tremendous size variation, the three smallest being in the 220–234 aa size range (all from members of the rhizobial group), and the three largest being in the 912–919 aa size range (all from members of the chlamydial group). Analyses of the DNA sequences of the encoding genes convinced us that these size assignments are essentially correct. Although one *P. aeruginosa* protein is of only 273 aas, and the *Myxococcus xanthus* protein is of 901 aas, almost all other homologues are in the 400–800 aa size range (Table 7).

A quick look at the organisms bearing secretin homologues reveals that all classes of proteobacteria as well as

Table 7
Recognized homologues of the Secretin family

Abbreviation	Database description	Organism	Bacillus type	Size	GI number
ComQ Asp	putative outer membrane protein ComQ	*Acinetobacter* sp.	γ3	723	12642803
RcpA Aac	Rough colony protein A	*Actinobacillus actinomycetemcomitans*	γ2	460	4768955
ExeD Ahy	general secretion pathway protein D precursor	*Aeromonas hydrophila*	γ1	678	1170050
Orf Ahy	S-protein secretion D	*Aeromonas hydrophila*	γ1	737	2126227
ExeD Asa	general secretion pathway protein D precursor	*Aeromonas salmonicida*	γ1	678	1170051
GspD Aae	general secretion pathway protein D	*Aquifex aeolicus*	Aqu	625	7514941
Orf Aae	conserved hypothetical protein	*Aquifex aeolicus*	Aqu	705	7514521
Orf fd	gene IV protein	Bacteriophage *fd* (*E. coli*)	γ1	426	9626336
Orf I2-2	gene IV protein	Bacteriophage *I2-2* (*E. coli*)	γ1	428	9625382
Orf If1	gene IV protein	Bacteriophage *If1* (*E. coli*)	γ1	429	9630755
Orf Ike	gene IV protein	Bacteriophage *Ike* (*E. coli*)	γ1	437	9626242
Orf M13	gene IV protein	Bacteriophage *M13* (*E. coli*)	γ1	426	138050
Orf Pf3	unknown protein	Bacteriophage *Pf3* (*Pseudomonas aeruginosa*)	γ3	430	9626321
RhcC Bja[a]	RhcC1	*Bradyrhizobium japonicum*	α	230	12620518
RhcC1 Bja	RhcC2	*Bradyrhizobium japonicum*	α	484	12620550
GspD Bce	GspD	*Burkholderia cepacia*	β	783	11559475
GspD Bps	general secretory pathway protein D	*Burkholderia pseudomallei*	β	750	4139236
Orf Cje	probable type II protein secretion system D protein	*Campylobacter jejuni*	ε	472	11347194
GspD Ccr	general secretion pathway protein D	*Caulobacter crescentus*	α	687	13421292
CpaC Ccr	CpaC	*Caulobacter crescentus*	α	560	7208425
SctC Cmu[b]	type III secretion protein	*Chlamydia muridarum*	Chla	918	11362809
GspD Cmu	general secretion pathway protein D	*Chlamydia muridarum*	Chla	759	11360973
YopC Ctr	secretion protein D	*Chlamydia trachomatis*	Chla	921	7469078
Orf Ctr	probable general secretion protein D	*Chlamydia trachomatis*	Chla	760	7468922
SctC Cpn	type III secretion protein	*Chlamydophila pneumoniae*	Chla	919	7468594
GspD Cpn	general secretion pathway protein D	*Chlamydophila pneumoniae*	Chla	754	7468239
Orf Cli	exporter protein	*Chlorobium limicola*	Chlo	461	10956078
GspD Dra	probable general secretion pathway protein D	*Deinococcus radiodurans*	Dei	740	7473495
HrcC Eam	HrcC	*Erwinia amylovora*	γ1	676	1336093
GspD Eco	probable general secretion pathway protein D	*Escherichia coli*	γ1	654	1170052
EtpD Eco	type II secretion pathway-related protein etpD	*Escherichia coli*	γ1	642	7466966
EivG Eco	type III secretion apparatus protein	*Escherichia coli*	γ1	567	12517375
HofQ Eco	protein transport protein HofQ precursor	*Escherichia coli*	γ1	412	1170332
EscC Eco	type III secretion system EscC protein	*Escherichia coli*	γ1	512	3414909
BfpB Eco	BfpB	*Escherichia coli*	γ1	552	1314252
PilN1 Eco	Lipoprotein	*Escherichia coli* plasmid ColIb-P9	γ1	560	9507539
PilN2 Eco	PilN	*Escherichia coli* plasmid R721	γ1	547	10955502
ComE Hin	competence protein E precursor	*Haemophilus influenzae*	γ2	445	1169008
PulD Kpn	general secretion pathway protein D precursor	*Klebsiella pneumoniae*	γ1	660	131592
LspD Lpn	type II outer membrane secretin	*Legionella pneumophila*	γ3	678	13625380
GspD Mlo	general secretion protein D	*Mesorhizobium loti*	α	708	13475694
Orf1 Mlo	pilus assembly protein	*Mesorhizobium loti*	α	481	13475417
Orf2 Mlo	type II secretion system protein	*Mesorhizobium loti*	α	432	13475294
CpaC Mlo	exporter protein	*Mesorhizobium loti*	α	471	13474660
Orf3 Mlo	probable secretory protein	*Mesorhizobium loti*	α	461	13471032
NolW Mlo[a]	nodulatin protein NolW	*Mesorhizobium loti*	α	220	13475297
PilQ Mxa	PilQ	*Myxococcus xanthus*	δ	901	3978519
PulD Ngo	outer membrane protein Omc precursor	*Neisseria gonorrhoeae*	β	711	548422
PilQ Ngo	PilQ protein	*Neisseria gonorrhoeae*	β	720	2120880
Orf1 Nme	pilus secretin	*Neisseria meningitidis*	β	761	11353851
Orf2 Nme	secretin precursor	*Neisseria meningitidis*	β	766	4027986
HrcC Pst	HrcC	*Pantoea stewartii*	γ1	677	9885640
ComE Pmu	ComE	*Pasteurella multocida*	γ2	444	12721580
RcpA Pmu	RcpA	*Pasteurella multocida*	γ2	470	12721161
OutD Pca	OutD protein	*Pectobacterium carotovorum*	γ1	649	479227
OutD1 Pch	general secretion pathway protein D precursor	*Pectobacterium chrysanthemi*	γ1	710	399825
OutD2 Pch	general secretion pathway protein D precursor	*Pectobacterium chrysanthemi*	γ1	712	399792
OutD3 Pch	general secretion pathway protein D precursor	*Pectobacterium chrysanthemi*	γ1	650	2506491
HrcC Pch	HrcC	*Pectobacterium chrysanthemi*	γ1	691	1772618
XcpQ1 Pae	general secretion pathway protein D precursor	*Pseudomonas aeruginosa*	γ3	658	544439

Table 7 (*continued*)

Abbreviation	Database description	Organism	Bacillus type	Size	GI number
XqhA Pae	Secretion protein XqhA	*Pseudomonas aeruginosa*	γ3	776	11352555
PilQ Pae	Fimbrial assembly protein PilQ precursor	*Pseudomonas aeruginosa*	γ3	714	12230952
Orf1 Pae	probable type II secretion system protein	*Pseudomonas aeruginosa*	γ3	759	11352405
PscC Pae	PscC	*Pseudomonas aeruginosa*	γ3	600	1781385
Orf2 Pae	probable type II secretion system protein	*Pseudomonas aeruginosa*	γ3	416	11352412
Orf3 Pae	probable type II secretion system protein	*Pseudomonas eruginosa*	γ3	803	11352402
Orf4 Pae	hypothetical protein	*Pseudomonas aeruginosa*	γ3	273	11349624
XcpQ2 Pal	outer membrane secretion protein Q	*Pseudomonas alcaligenes*	γ3	649	3978475
XcpQ Ppu	protein secretion protein xcpQ precursor	*Pseudomonas putida*	γ3	591	2120685
HrcC Psy	HrcC	*Pseudomonas syringae*	γ3	700	3228547
HrpH Psy	hypersensitivity response secretion protein HrpH precursor	*Pseudomonas syringae*	γ3	701	6016255
HrpA Rso	hypersensitivity response secretion protein	*Ralstonia solanacearum*	β	568	2833448
NolW Rfr[a]	nodulation protein NolW	*Rhizobium fredii* (*Sinorhizobium fredii*)	α	234	462733
Orf Rsp	hypothetical 44.3 kDa protein	*Rhizobium* sp.	α	423	2495099
PilN Sty[c]	PilNa	*Salmonella typhi*	γ1	553	7274588
InvG1 Sty	InvG protein precursor	*Salmonella typhimurium*	γ1	562	1170574
InvG2 Sty	invasion protein invG	*Salmonella typhimurium*	γ1	563	2126157
SpiA Sty	SpiA	*Salmonella typhimurium*	γ1	497	1498307
MxiD Sfl	outer membrane protein MxiD precursor	*Shigella flexneri*	γ1	566	13449092
MxiD Sso	outer membrane protein MxiD precursor	*Shigella sonnei*	γ1	566	2495097
GspD Ssp	general secretion pathway protein D	*Synechocystis* sp.	Cya	785	7469324
ComE Tma	hypothetical protein	*Thermotoga maritima*	The	1285	7462739
GspD Tma	general secretion pathway protein D	*Thermotoga maritima*	The	387	7462809
EspD Vch	general secretion pathway protein D precursor	*Vibrio cholerae*	γ1	674	11182423
Orf1 Vch	fimbrial assembly protein	*Vibrio cholerae*	γ1	578	11354911
MshD Vch	mannose-sensitive hemagglutinin D	*Vibrio cholerae*	γ1	559	791156
Orf2 Vch	similar to gene IV protein	*Vibrio cholerae*	γ1	500	9630770
TcpC Vch	toxin corregulated pilus biosynthesis outer membrane protein C	*Vibrio cholerae*	γ1	489	267086
XpsD Xca	general secretion pathway protein D precursor	*Xanthomonas campestris*	γ4	759	129748
HrpA Xca	hypersensitivity response secretion protein	*Xanthomonas campestris*	γ4	607	462304
HrpA Xor	HrpA	*Xanthomonas oryzae*	γ4	605	7350909
Orf1 Xfa	general secretory pathway protein D precursor	*Xylella fastidiosa*	γ4	775	11360974
Orf2 Xfa	fimbrial assembly protein	*Xylella fastidiosa*	γ4	637	11360960
YscC Yet	Yop proteins translocation protein C precursor	*Yersinia enterocolitica*	γ1	607	10955572
YsaC Yet	YsaC	*Yersinia enterocolitica*	γ1	525	8996028
Ysc Ype	Yop proteins translocation protein C homolog	*Yersinia pestis*	γ1	607	10955619

[a] These three nodulation proteins were excluded from the phylogenetic tree because of their small sizes and lack of homology with the most highly conserved domains in all other secretin.

[b] The database entry for this protein indicated a size of 672 aas. However, comparison with the *C. trachomatis* and *C. pneumoniae* homologues revealed that the initiation codon had been incorrectly assigned. The proposed length for this protein in 918 aas and the initiation codon encodes a valine.

[c] The database entry for this protein indicates a size of 277aas. The nucleotide sequence was translated in all three reading frames, and the translated sequences were tested for homology and accuracy by using the BlastX program. The protein used in our analysis is 553 aas long.

several others (*Aquifex*, *Chlamydia*, *Chlorobium*, *Deinococcus*, *Synechocystis* and *Thermatoga*) are represented. Of divergent bacteria with fully sequenced genomes, only *Deinococcus* and *Synechocystis* have a single secretin while *A. aeolicus*, all of the chlamydial species, and *T. maritima* have two paralogues. It is interesting to note that all phage-encoded secretins are in a single size range (420–437 aas), while the two plasmid-encoded proteins mentioned above are similarly of about the same size (560 and 547 aa, respectively). As will be discussed below, size differences correlate with phylogenetic grouping and functional type (Table 8).

In constructing the multiple alignment, three sequences were omitted. These proteins were the three short proteins, NolW Mlo, NolW Rfr and RhcC Bja. They aligned at alignment positions 280–608 in the multiple alignment presented on our ALIGN web site. These three proteins lack the most conserved regions of all other secretins. Examination of the genes encoding these proteins led us to conclude that the size assignments are essentially correct. Although these proteins are believed to function in nodulation, their biochemical functions are unknown.

Examination of the multiple alignment of the remaining 93 secretins revealed that no residue is fully conserved. However, one G is conserved in all but one protein (alignment position 1123; see below). The major size differences between the various secretins occur in the N-terminal 800 residue positions which are not present in all homologues. However, residue positions 813–1171 are represented in every homologue except the three Rhizobial proteins mentioned above, and this short region proved to be the best conserved (see our ALIGN web site). Nine or ten well-conserved probable amphipathic β-strands were identified within this region.

Three extended, well-conserved, consensus sequences were derived from the three most conserved regions as follows:

1. $(L\ I\ V)_2\ X_2\ L\ D\ (L\ I\ V)\ (P\ A\ R)\ X_2\ Q\ V\ X\ (L\ I\ V)\ E\ A\ X\ (L\ I\ V)_2\ E$ (alignment positions 813–833).
2. $(L\ I\ V\ F\ W)\ X\ (L\ I\ V\ M\ A)\ X\ (L\ I\ V)\ (S\ T\ N\ Q)\ A\ L\ X_7\ (L\ I\ V)_2\ (S\ T)\ X\ P\ X\ (L\ I\ V)_2\ T$ (alignment positions 972–995).
3. $(L\ I\ V)\ X\ (L\ I\ V)\ X_2\ G\ X\ (S\ T)\ (L\ I\ V)_3\ G\ G^*\ (L\ I\ V)_2\ X_{10}\ V\ P\ (L)_2\ (G\ S)\ D\ (L\ I\ V)\ P\ (L\ I\ V)_2\ G\ X\ L\ F\ R\ X_{10}\ (L\ I\ V\ M\ F)_5\ X\ P\ X\ (L\ I\ V)_2$ (alignment positions 1111–1171).

The most conserved residue, the G* in consensus sequence #3 is conserved in all proteins except TcpC Vch (see Table 6).

The phylogenetic tree for the Secretin family is shown in Fig. 5A, and the corresponding 16S rRNA tree is shown in Fig. 5B. Ten clusters of secretins as well as 11 branches bearing a single protein were observed. Two of these clusters (clusters 4 and 8) bear only chlamydial proteins. One cluster (cluster 10) bears all of the phage proteins. Finally, all remaining clusters bear only proteobacterial proteins with the exception of cluster 5, which also bears a sequence divergent protein from the green bacterium *Chlorobium*. All other proteins from bacteria outside of the proteobacterial group are found singly on branches that lack near homologues. Thus, there is no evidence for horizontal transfer of genes encoding secretins between the major bacterial kingdoms.

Table 8 summarizes much of the phylogenetic data presented in Fig. 5A. Comparing Fig. 5A and B, clusters 1–3, 5–7 and 9 include proteobacterial proteins that in general do not follow the phylogenies of the organisms. Moreover, with the exception of cluster 9, single organisms bear multiple paralogues within each of these clusters. These facts must reflect late gene duplication events and/or horizontal transfer of genes between proteobacteria. They show that the proteins in these clusters do not exhibit simple orthologous relationships. Only the two chlamydial clusters (clusters 4 and 8) do exhibit the expected orthologous relationships. We can presume that all three proteins in each of these chlamydial clusters serve the same function.

The third column in Table 8 provides the average sizes of the proteins in the various clusters ± S.D., while column 4 indicates their presumed functions. It can be seen that each phylogenetic cluster exhibits a distinctive size that may correlate with function. Thus,

(1) Four clusters (1, 2, 4 and 9) include proteins in the size range 665–758 aas, and these secretins are probably all constituents of IISP systems (C. Peabody, M.-R. Yen and M.H. Saier, Jr., unpublished—in prep.). Two outliers, GspD Aae and LspD Lpn, adjacent to cluster 9 in Fig. 5A, are also constituents of IISP systems.

(2) Cluster 3 consists of two subclusters with very different sizes and functions. Cluster 3a includes proteins of 412–445 aas which function in competence (DNA

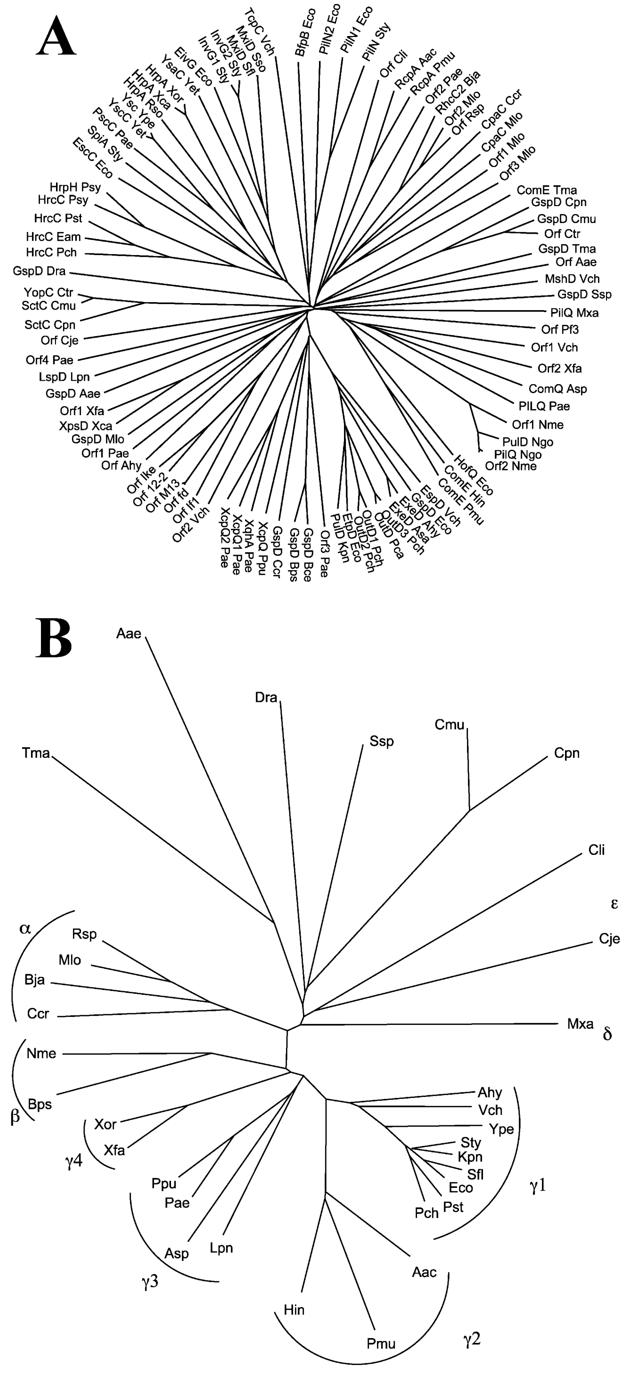

Fig. 5. The secretin family. (A) Phylogenetic tree for currently recognized members of the Secretin family. (B) 16S rRNA tree for organisms known to possess one or more Secretin family homologue(s). Protein abbreviations are as indicated in Table 7. The three nodulation proteins (footnote a in Table 7) were omitted due to their small size.

Table 8
Organismal representation, size variation, and functional assignments for the phylogenetic clusters of secretins

Cluster[a]	Organisms represented[b]	Average size ± S.D.[c]	(Putative) function[d]
1	γ3, α, β	712 ± 77	TIISP
2	γ1	665 ± 30	TIISP
3a	γ1, γ3, γ4, β	701 ± 63	pilus assembly
3b	γ1, γ2	434 ± 19	competence
4	*Chlamydia*	758 ± 3	TIISP
5	γ2, γ3, α, (*Chlorobium*)	442 ± 80	rough colony phenotype (fimbrium assembly)
6	γ1	540 ± 29	pilus assembly
7	γ1, γ3, γ4, β	606 ± 67	TIIISP
8	*Chlamydia*	919 ± 2	TIIISP
9	γ1, γ4, α	748 ± 26	TIISP
10	γ1 (phage)	441 ± 29	phage assembly
11[e]	α (*Rhizobia*)	228 ± 7	nodulation

[a] Cluster refers to the phylogenetic cluster shown in Fig. 5A.

[b] All Greek letter entries refer to the category of proteobacteria from which the proteins are derived.

[c] Sizes are expressed in numbers of amino acid residues ± S.D.

[d] TIISP and TIIISP refer to type II secretory pathway (main terminal branch) and type III secretory pathway (pathogenesis-related systems), respectively.

[e] The three nodulation proteins from Rhizobial species are not included in Fig. 5A, but they comprise a unique cluster.

uptake), while cluster 3b includes proteins of 578–766 aas, which affect colony morphology and may function in pilus assembly [87]. The phylogenetic clustering of these two subclusters might be interpreted to suggest a role of pili in competence.

(3) In addition to cluster 3b, clusters 5 and 6 are probably concerned with pilus assembly [10]. The size ranges of these two clusters differ significantly from each other (442 ± 80 versus 540 ± 29 aas). That of the cluster 3b secretins is 434 ± 19 aas, the same as that of cluster 5. As there are many distinct types of pili (fimbriae), it can be suggested that each of these three clusters is concerned with the biogenesis of a different type of pilus.

(4) Cluster 7 and 8 secretins are constituents of IIISP systems [10]. The chlamydial cluster 8 proteins are about 300 residues longer than the cluster 7 secretins. This proved to be due to the presence in the former proteins of long N-terminal extensions that lack sequence similarity with anything else in the databases.

(5) Cluster 10 proteins are phage-encoded (plus one protein from *V. cholerae*). These proteins, of uniform size (441 ± 29 aas), are presumed to function in phage particle export and assembly. They are in the same size range as most of the pilin export secretins, possibly suggesting that the phage proteins were derived from the latter.

(6) Finally, cluster 11 rhizobial proteins (not presented in Fig. 5A) are all exceptionally small (228 ± 7 aas), lacking the best-conserved portions of all other secretins. While these proteins function in nodulation, their specific biochemical roles are not known.

8. Conclusions

Table 9 summarizes the properties and organismal distributions of the five types of Gram-negative bacterial outer membrane porins that mediate protein secretion analyzed in this report. All of these families include roughly (within a two-fold range) the same numbers of sequenced members at the time when this work was completed (July 2001). Thus, the smallest family (FUP) has 58 members, while the largest family (AT) has slightly more than twice this number, 120 members. With respect to size variation, the FUP family members show a restricted size range (843 ± 36) as does the OMF family (462 ± 32), but the size variance of the three remaining families is much greater, particularly for the AT family (1261 ± 589). The TPS and Secretin families show nearly the same average size with nearly the same standard deviation value (649 ± 140 as compared with 619 ± 149). In all five families, N-terminal hydrophobic leader sequences can be found that presumably target all of these cytoplasmically synthesized proteins to the GSP (IISP) secretory apparatus for export across the inner (cytoplasmic) membrane. In the periplasm, the targeting sequences are

Table 9
Properties and organismal distribution of outer membrane porins mediating protein secretion

Family	FUP	OMF	AT	TPS	Secretin
TC #	(1.B.11)	(1.B.17)	(1.B.12)	(1.B.20)	(1.B.22)
# Homologues	58	102	120	77	96
Size range	729–895	347–541	418–3705	398–1056	273–1285
Average size ± S.D.	843 ± 36	462 ± 32	1261 ± 589	649 ± 140	619 ± 149
Leader sequence	+	+	+	+	+
Putative #β-strands	24	8	14	19	12
# Homologues/organelle type					
Proteobacteria					
γ1	48	21	43	11	38
γ2	1	4	8	13	4
γ3	4	31	9	8	15
γ4	1	3	4	4	5
β	1	10	17	8	7
α	1	14	26	9	12
ε		7	6	3	1
δ					1
Chlamydiales			7	3	6
Deinococcus	1	1		1	1
Cyanobacteria	1	1		2	1
Spirochaetales		1			
Aquificales		6		2	2
Porphyromonas		3			
Chlorobium					1
Thermotogales					2
Fusobacteria				1	
Thermotogales				1	
Eukaryotes					
Animal				3	
Plant				6	
Fungi				2	

removed, and the proteins fold, either in the periplasm, or on the periplasmic surface of the outer membrane in preparation for insertion into the outer membrane where they function in the export of target proteins.

Each of the five families of OMPs functions by a distinct mechanism, often in conjunction with different sets of auxiliary proteins. Thus, fimbrial ushers function together with periplasmic chaperone proteins, and these two proteins cooperate to fold the fimbrial structural subunits before export to the cell surface in preparation for assembly of the fimbrium. By contrast, the OMFs function together with a primary cytoplasmic membrane transporter and a membrane fusion protein (see Introduction) to export an extended, unfolded protein which presumably folds to its native configuration only after it has crossed the outer membrane via the OMF pore. This situation contrasts with the ATs, which usually (but probably not always) export their own N-terminal domains, which can then be processed and folded on the external side. Often (but not always) the C-terminal AT domain is proteolytically cleaved from the exported virulence factor on the external surface. Finally, while the secretins, functioning with type II main terminal branch (MTB)-type secretory systems or with the type III pathogenicity-related systems, are believed to always export fully folded proteins of varied structures and functions, the TPS-type systems each usually exports a highly specific protein substrate that is encoded within a single operon with the TPS-type exporter. In a recent review, Jacob-Dubuisson et al. [19] have designated the substrate protein as "TpsA" and the transporter as "TpsB." While exceptions undoubtedly will prove to exist, the occurrence of simple two-partner systems appears to be the general rule. The degree to which these systems will prove capable of interchanging their substrates (i.e., see Ref. [66]) remains to be determined. In many of these systems, the targeting sequences are still poorly defined, and consequently the molecular basis for specificity is not well understood.

As summarized in Table 9, the organismal distributions of the five outer membrane protein secretion channel-types vary widely. All of them have been identified in all of the major proteobacterial subdivisions with the exception of FUPs, which have not yet been found in the δ- or ε-proteobacteria, and the OMFs, ATs and TPSs that have not yet been identified in the δ-subdivision. It should be noted, however, that only two ε-proteobacterial genomes (those of *H. pylori* and *C. jejuni*) have been fully sequenced, and no fully sequenced δ-proteobacterial genome is as yet available for analysis.

Outside of the proteobacteria, the ATs are most restricted in distribution, being found only in the chlamydial kingdom, while the FUPs have been identified only in cyanobacteria and in *Deinococcus*. The OMFs have not been identified in several bacterial kingdoms, but they are present in five divergent bacterial kingdoms, those including *Deinococcus*, *Aquifex*, *Porphyromonas*, the cyanobacterium, *Synechocystis*, and the spirochete, *Borrelia burgdorferi*. Finally, homologues of TPS-type systems are found not only in four non-proteobacterial kingdoms, they are also found in three eukaryotic kingdoms where one such protein, Oep-75 of the pea, has been shown to be a component of a chloroplast protein import system [74,76]. Thus, while the TPS family members function in protein export in bacteria, they function in protein import in chloroplasts, exhibiting apparent reverse polarity. None of the families studied was found in Gram-positive bacteria or archaea with the sole exception of *D. radiodurans*, a dual membrane organism that has erroneously been assigned to the Gram-positive bacterial kingdom [39,51].

In several cases, the phylogenetic analyses led to specific functional predictions. Thus, by comparing the protein phylogenetic trees with corresponding 16S rRNA trees, we were able to provide evidence for orthologous relationships where all putative orthologues within a cluster are presumed to serve a unified function. A striking example is the TPS family, where three large clusters of proteobacterial proteins were found (Fig. 4A). Each of these clusters included members, almost all of which were of uniform size. Moreover, in two of these proteobacterial clusters (clusters 3 and 5 in Fig. 4A), the phylogenies of the proteins corresponded to those of the 16S rRNAs within experimental error, and no organism exhibited more than a single member. By contrast, in the third proteobacterial cluster (cluster 6 in Fig. 4A) the phylogenies of the proteins did not follow those of the 16S rRNAs, and several organisms were found to have multiple paralogues within the cluster. Thus, both late gene duplication events and horizontal gene transfer may account for the anomalous features of this third phylogenetic cluster.

The analyses reported here illustrate the utility of the phylogenetic approach for (1) making functional predictions, (2) understanding the evolutionary pathways taken for the dissemination of members of a family in distantly related living organisms, and (3) identifying conserved structural and sequence features that serve to characterize a particular protein family. The expansion of available software for more detailed analyses of the type reported here should greatly enhance such endeavors.

Acknowledgements

Work in our laboratory was supported by NIH grants GM55434 and GM64368 from the National Institute of General Medical Sciences. We thank Mary Beth Hiller for her assistance in the preparation of this manuscript.

References

[1] M.H. Saier Jr., A functional–phylogenetic classification system for transmembrane solute transporters, Microbiol. Mol. Biol. Rev. 64 (2000) 354–411.

[2] M. Sandkvist, Biology of type II secretion, Mol. Microbiol. 40 (2001) 271–283.

[3] J.-W. de Gier, J. Luirink, Biogenesis of inner membrane proteins in *Escherichia coli*, Mol. Microbiol. 40 (2001) 314–322.

[4] J. Luirink, T. Samuelsson, J.-W. de Gier, YidC/Oxa1p/Alb3: evolutionarily conserved mediators of membrane protein assembly, FEBS Lett. 501 (2001) 1–5.

[5] M.-R. Yen, K.T. Harley, Y.-H. Tseng, M.H. Saier Jr., Phylogenetic and structural analyses of the Oxa1 family of protein translocases, FEMS Microbiol. Lett. 204 (2001) 223–231.

[6] M.H. Saier Jr., Families of transmembrane transporters selective for amino acids and their derivatives, Microbiology 146 (2000) 1775–1795.

[7] R. Voulhoux, G. Ball, B. Ize, M.L. Vasil, A. Lazdunski, L.-F. Wu, A. Filloux, Involvement of the twin-arginine translocation system in protein secretion via the type II pathway, EMBO J. 20 (2001) 6735–6741.

[8] T.B. Cao, M.H. Saier Jr., Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions, Microbiology 147 (2001) 3201–3214.

[9] P.J. Christie, Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines, Mol. Microbiol. 40 (2001) 294–305.

[10] L. Nguyen, I.T. Paulsen, J. Tchieu, C.J. Hueck, M.H. Saier Jr., Phylogenetic analyses of the constituents of type III protein secretion systems, J. Mol. Microbiol. Biotechnol. 2 (2000) 125–144.

[11] G.V. Plano, J.B. Day, F. Ferracci, Type III export: new uses for an old pathway, Mol. Microbiol. 40 (2001) 284–293.

[12] O. Mol, B. Oudega, Molecular and structural aspects of fimbriae biosynthesis and assembly in *Escherichia coli*, FEMS Microbiol. Rev. 19 (1996) 25–52.

[13] M. Van Rosmalen, M.H. Saier Jr., Structural and evolutionary relationships between two families of bacterial extracytoplasmic chaperone proteins which function cooperatively in fimbrial assembly, Res. Microbiol. 144 (1993) 507–527.

[14] H. Wu, P.M. Fives-Taylor, Molecular strategies for fimbrial expression and assembly, Crit. Rev. Oral Biol. Med. 12 (2001) 101–115.

[15] R. Binet, S. Létoffé, J.M. Ghigo, P. Delepelaire, C. Wandersman, Protein secretion by Gram-negative bacterial ABC exporters—a review, Gene 192 (1997) 7–11.

[16] L.A. Fernandez, V. de Lorenzo, Formation of disulphide bonds during secretion of proteins through the periplasmic-independent type I pathway, Mol. Microbiol. 40 (2001) 332–346.

[17] I.T. Paulsen, J.H. Park, P.S. Choi, M.H. Saier Jr., A family of Gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from Gram-negative bacteria, FEMS Microbiol. Lett. 156 (1997) 1–8.

[18] B.J. Loveless, M.H. Saier Jr., A novel family of autotransporting, channel-forming, bacterial virulence proteins, Mol. Membr. Biol. 14 (1997) 113–123.

[19] F. Jacob-Dubuisson, C. Locht, R. Antoine, Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins, Mol. Microbiol. 40 (2001) 306–313.

[20] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[21] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.

[22] D.-F. Feng, R.F. Doolittle, Progressive alignment and phylogenetic tree construction of protein sequences, Methods Enzymol. 183 (1990) 375–387.

[23] R.D.M. Page, TREEVIEW: an application to display phylogenetic trees on personal computers, Comput. Appl. Biosci. 12 (1996) 357–358.

[24] G.B. Young, D.L. Jack, D.W. Smith, M.H. Saier Jr., The amino acid/auxin:proton symport permease family, Biochim. Biophys. Acta 1415 (1999) 306–322.

[25] K. Hofmann, W. Stoffel, Tmbase—a database of membrane spanning protein segments, Biol. Chem. 347 (1993) 166.

[26] G. von Heijne, Membrane protein structure prediction, hydrophobicity analysis and the positive-inside rule, J. Mol. Biol. 225 (1992) 487–494.

[27] Y. Zhai, M.H. Saier Jr., A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence, J. Mol. Microbiol. Biotechnol. 3 (2001) 501–502.

[28] Y. Zhai, M.H. Saier Jr., A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins, J. Mol. Microbiol. Biotechnol. 3 (2001) 285–286.

[29] J.A. Cuff, M.E. Clamp, A.S. Siddiqui, M. Finlay, G.J. Barton, Jpred: a consensus secondary structure prediction server, Bioinformatics 14 (1998) 892–893.

[30] D.T. Jones, W.R. Taylor, J.M. Thornton, A model recognition approach to the prediction of all-helical membrane protein structure and topology, Biochemistry 33 (1994) 3038–3049.

[31] T. Le, T.T. Tseng, M.H. Saier Jr., Flexible programs for the prediction of average amphipathicity of multiply aligned homologous proteins: application to integral membrane transport proteins, Mol. Membr. Biol. 16 (1999) 173–179.

[32] K.W. Dodson, F. Jacob-Dubuisson, R.T. Striker, S.J. Hultgren, Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 3670–3674.

[33] I. Vallet, J.W. Olson, S. Lory, A. Lazdunski, A. Filloux, The chaperone/usher pathways of *Pseudomonas aeruginosa*: identification of fimbrial gene clusters (cup) and their involvement in biofilm formation, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 6911–6916.

[34] O. Mol, W.C. Oudhuis, R.P. Oud, R. Sijbrandi, J. Luirink, N. Harms, B. Oudega, Biosynthesis of K88 fimbriae in *Escherichia coli*: interaction of tip-subunit FaeC with the periplasmic chaperone FaeE and the outer membrane usher FaeD, J. Mol. Microbiol. Biotechnol. 3 (2001) 135–142.

[35] Q.A. Valent, J. Zaal, F.K. de Graaf, B. Oudega, Subcellular localization and topology of the K88 usher FaeD in *Escherichia coli*, Mol. Microbiol. 16 (1995) 1243–1257.

[36] J. Cao, A.S. Khan, M.E. Bayer, D.M. Schifferli, Ordered translocation of 987P fimbrial subunits through the outer membrane of *Escherichia coli*, J. Bacteriol. 177 (1995) 3704–3713.

[37] C.J. Smyth, M.B. Marron, J.M. Twohig, S.G. Smith, Fimbrial adhesins: similarities and variations in structure and biogenesis, FEMS Immunol. Med. Microbiol. 16 (1996) 127–139.

[38] D.G. Thanassi, E.T. Saulino, M.-J. Lombardo, R. Roth, J. Heuser, S.J. Hultgren, The PapC usher forms an oligomeric channel: implications for pilus biogenesis across the outer membrane, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 3146–3151.

[39] O. White, J.A. Eisen, J.F. Heidelberg, E.K. Hickey, J.D. Peterson, R.J. Dodson, D.H. Haft, M.L. Gwinn, W.C. Nelson, D.L. Richardson, K.S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J.J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K.S. Makarova, L. Aravind, M.J. Daly, K.W. Minton, R.D. Ketchum, K.A. Ketchum, K.E. Nelson, S. Salzberg, H.O. Smith, J.C. Venter, C.M. Fraser, Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1, Science 286 (1999) 1571–1577.

[40] D. Jeanteur, J.H. Lakey, F. Pattus, The bacterial porin superfamily: sequence alignment and structure prediction, Mol. Microbiol. 5 (1991) 2153–2164.

[41] S.S. Pao, I.T. Paulsen, M.H. Saier Jr., Major facilitator superfamily, Microbiol. Mol. Biol. Rev. 62 (1998) 1–34.

[42] W. Saurin, M. Hofnung, E. Dassa, Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters, J. Mol. Evol. 48 (1999) 22–41.

[43] T.-T. Tseng, K.S. Gratwick, J. Kollman, D. Park, D.H. Nies, A. Goffeau, M.H. Saier Jr., The RND permease superfamily: an ancient,

ubiquitous and diverse family that includes human disease and development proteins, J. Mol. Microbiol. Biotechnol. 1 (1999) 107–125.

[44] K.T. Harley, M.H. Saier Jr., A novel ubiquitous family of putative efflux transporters, J. Mol. Microbiol. Biotechnol. 2 (2000) 195–198.

[45] T. Dinh, I.T. Paulsen, M.H. Saier Jr., A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of Gram-negative bacteria, J. Bacteriol. 176 (1994) 3825–3831.

[46] V. Koronakis, J. Li, E. Koronakis, K. Stauffer, Structure of TolC, the outer membrane component of the bacterial type I efflux system, derived from two-dimensional crystals, Mol. Microbiol. 23 (1997) 617–626.

[47] S. Létoffé, P. Delepelaire, C. Wandersman, Protein secretion in Gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding, EMBO J. 15 (1996) 5804–5811.

[48] V. Koronakis, A. Sharff, E. Koronakis, B. Luisi, C. Hughes, Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export, Nature 405 (2000) 914–919.

[49] X.Z. Li, K. Poole, Mutational analysis of the OprM outer membrane component of the MexA–MexB–OprM multidrug efflux system of *Pseudomonas aeruginosa*, J. Bacteriol. 183 (2001) 12–27.

[50] K.K. Wong, F.S. Brinkman, R.S. Benz, R.E. Hancock, Evaluation of a structural model of *Pseudomonas aeruginosa* outer membrane protein OprM, an efflux component involved in intrinsic antibiotic resistance, J. Bacteriol. 183 (2001) 367–374.

[51] G.J. Olsen, C.R. Woese, R. Overbeek, The winds of (evolutionary) change: breathing new life into microbiology, J. Bacteriol. 176 (1994) 1–6.

[52] P. Benjelloun-Touimi, J. Sansonetti, C. Parsot, SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion, Mol. Microbiol. 17 (1995) 123–135.

[53] T.M. Finn, L.A. Stevens, Tracheal colonization factor: a *Bordetella pertussis* secreted virulence determinant, Mol. Microbiol. 16 (1995) 625–634.

[54] J. Jose, F. Jãhnig, T.F. Meyer, Common structural features of IgA1 protease-like outer membrane protein autotransporters, Mol. Microbiol. 18 (1995) 377–382.

[55] T. Suzuki, M.C. Lett, C. Saskawa, Extracellular transport of VirG protein in *Shigella*, J. Biol. Chem. 270 (1995) 30874–30880.

[56] I.R. Henderson, R. Cappello, J.P. Nataro, Autotransporter proteins, evolution and redefining protein secretion, Trends Microbiol. 8 (2000) 529–532.

[57] I.R. Henderson, J.P. Nataro, Virulence functions of autotransporter proteins, Infect. Immun. 69 (2001) 1231–1243.

[58] J. Maurer, J. Jose, T.F. Meyer, Characterization of the essential transport function of the AIDA-I autotransporter and evidence supporting structural predictions, J. Bacteriol. 181 (1999) 7014–7020.

[59] D.M. Guyer, I.R. Henderson, J.P. Nataro, H.L.T. Mobley, Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*, Mol. Microbiol. 38 (2000) 53–66.

[60] J.L. Shannon, R.C. Fernandez, The C-terminal domain of the *Bordetella pertussis* autotransporter BrkA forms a pore in lipid bilayer membranes, J. Bacteriol. 181 (1999) 5838–5842.

[61] J.W. St. Geme III, D. Cutter, The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C-terminus and fully cell associated, J. Bacteriol. 182 (2000) 6005–6013.

[62] C.T. Lattemann, J. Maurer, E. Gerland, T.F. Meyer, Autodisplay: functional display of active β-lactamase on the surface of *Escherichia coli* by the AIDA-I autotransporter, J. Bacteriol. 182 (2000) 3726–3733.

[63] K. Poole, E. Schiebel, V. Braun, Molecular characterization of the hemolysin determinant of *Serratia marcescens*, J. Bacteriol. 170 (1988) 3177–3188.

[64] R. Hertle, S. Brutsche, W. Groeger, S. Hobbie, W. Kock, U. Könninger, V. Braun, Specific phosphatidylethanolamine dependence of *Serratia marcescens* cytotoxin activity, Mol. Microbiol. 26 (1997) 853–865.

[65] I. Hirono, N. Tange, T. Aoki, Iron-regulated haemolysin gene from *Edwardsiell tarda*, Mol. Microbiol. 24 (1997) 851–856.

[66] F. Jacob-Dubuisson, C. Buisine, E. Willery, G. Renauld-Mongénie, C. Locht, Lack of fundamental complementation between *Bordetella pertussis* filamentous hemagglutinin and *Proteus mirabilis* HpmA hemolysin secretion machineries, J. Bacteriol. 179 (1997) 775–783.

[67] F. Jacob-Dubuisson, B. Kehoe, E. Willery, N. Reveneau, C. Locht, D.A. Relman, Molecular characterization of *Bordetella bronchiseptica* filamentous haemagglutinin and its secretion machinery, Microbiology 146 (2000) 1211–1221.

[68] K.L. Palmer, R.S. Munson Jr., Cloning and characterization of the genes encoding the hemolysin of *Haemophilus ducreyi*, Mol. Microbiol. 18 (1995) 821–830.

[69] F. Jacob-Dubuisson, C. El-Hamel, N. Saint, S. Guèdin, E. Willery, G. Molle, C. Locht, Channel formation by FhaC, the outer membrane protein involved in the secretion of the *Bordetella pertussis* filamentous hemagglutinin, J. Biol. Chem. 274 (1999) 37731–37735.

[70] U.W. Könninger, S. Hobbie, R. Benz, V. Braun, The haemolysin-secreting ShlB protein of the outer membrane of *Serratia marcescens*: determination of surface-exposed residues and formation of ion-permeable pores by ShlB mutants in artificial lipid bilayer membranes, Mol. Microbiol. 32 (1999) 1212–1225.

[71] S. Guédin, E. Willery, J. Tommassen, E. Fort, H. Drobecq, C. Locht, F. Jacob-Dubuisson, Novel topological features of FhaC, the outer membrane transporter involved in the secretion of the *Bordetella pertussis* filamentous hemagglutinin, J. Biol. Chem. 275 (2000) 30202–30210.

[72] M. Akita, E. Nielsen, K. Keegstra, Identification of protein transport complexees in the chloroplastic envelope membranes via chemical cross-linking, J. Cell Biol. 136 (1997) 983–994.

[73] E. Nielsen, M. Akita, J. Davila-Aponte, K. Keegstra, Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone, EMBO J. 16 (1997) 935–946.

[74] D.J. Schnell, F. Kessler, G. Blobel, Isolation of components of the chloroplast protein import machinery, Science 266 (1994) 1007–1012.

[75] P.J. Tranel, K. Keegstra, A novel, bipartite transit peptide targets OEP75 to the outer membrane of the chloroplastic envelope, Plant Cell 8 (1996) 2093–2104.

[76] P.J. Tranel, J. Froehlich, G. Goyal, K. Keegstra, A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway, EMBO J. 14 (1995) 2436–2446.

[77] S.C. Hinnah, K. Hill, R. Wagner, T. Schlicher, J. Soll, Reconstitution of a chloroplast protein import channel, EMBO J. 16 (1997) 7351–7360.

[78] W. Bitter, M. Koster, M. Latijnhouwers, H. de Cock, J. Tommassen, Formation of oligomeric rings by XcpQ and PilQ, which are involved in protein transport across the outer membrane of *Pseudomonas aeruginosa*, Mol. Microbiol. 27 (1998) 209–219.

[79] G.R. Cornelis, A. Boland, A.P. Boyd, C. Geuijen, M. Iriarte, C. Neyt, M.-P. Sory, I. Stainier, The virulence plasmid of *Yersinia*, and antihost genome, Microbiol. Mol. Biol. Rev. 62 (1998) 1315–1352.

[80] N.-T. Hu, M.-N. Hung, D.C. Chen, R.-T. Tsai, Insertion mutagenesis of XpsD, an outer-membrane protein involved in extracellular protein secretion in *Xanthomonas campestris* pv. *campestris*, Microbiology 144 (1998) 1479–1486.

[81] R.F. Collins, L. Davidsen, J.P. Derrick, R.C. Ford, T. Tønjum, Analysis of the PilQ secretin from *Neisseria meningitidis* by transmission electron microscopy reveals a dodecameric quaternary structure, J. Bacteriol. 183 (2001) 3825–3832.

[82] R. Brok, P. Van Gelder, M. Winterhalter, U. Ziese, A.J. Koster, H. de Cock, M. Koster, J. Tommassen, W. Bitter, The C-terminal domain of the *Pseudomonas* secretin XcpQ forms oligomeric rings with pore activity, J. Mol. Biol. 294 (1999) 1169–1179.

[83] A. Martínez, P. Ostrovsky, D.N. Nunn, Identification of an additional member of the secretin superfamily of proteins in *Pseudomonas aeruginosa* that is able to function in type II protein secretion, Mol. Microbiol. 28 (1998) 1235–1246.

[84] B.M. Davis, E.H. Lawson, M. Sandkvist, A. Ali, S. Sozhamannan, M. Waldor, Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXϕ, Science 288 (2000) 333–335.

[85] D.K. Marciano, M. Russel, S.M. Simon, An aqueous channel for filamentous phage export, Science 284 (1999) 1516–1519.

[86] S.A. Schmidt, D. Bieber, S.W. Ramer, J. Hwang, C.-Y. Wu, G. Schoolnik, Structure-function analysis of BfpB, a secretin-like protein encoded by the bundle-forming-pilus operon of enteropathogenic *Escherichia coli*, J. Bacteriol. 183 (2001) 4848–4859.

[87] E.M. Haase, J.L. Zmuda, F.A. Scannapieco, Identification and molecular analysis of rough-colony-specific outer membrane proteins of *Actinobacillus actinomycetemcomitans*, Infect. Immun. 67 (1999) 2901–2908.