J. Mol. Microbiol. Biotechnol. (2000) 2(2): 125-144.

Phylogenetic Analyses of the Constituents of Type III Protein Secretion Systems

Lily Nguyen[1], Ian T. Paulsen[2], Jason Tchieu, Christoph J. Hueck[3], and Milton H. Saier, Jr.*

Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA
[1]Present Address: Centers for Disease Control and Prevention, 1600 Clifton Road NE, MS G-19, Atlanta, GA 30333, USA
[2]Present Address: The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 20850, USA
[3]CREATOGEN GmbH, Ulmer Strasse 160a, 86156 Augsburg, Germany

Abstract

Multicomponent Type III protein secretion systems transfer Gram-negative bacterial virulence factors directly from the bacterial cytoplasm to the cytoplasm of a host eukaryotic cell in a process that may involve a single energy-coupled step. Extensive evidence supports the conclusion that the genetic apparatuses that encode these systems have been acquired independently by different Gram-negative bacteria, presumably by lateral transfer. In this paper we conduct phylogenetic analyses of currently sequenced constituents of these systems and their homologues. The results reveal the relative relatedness of these systems and show that they evolved with little or no exchange of constituents between systems. This fact suggests that horizontal transmission of the genes encoding these systems always occurred as a unit without the formation of hybrid gene clusters. Moreover, homologous flagellar proteins show phylogenetic clustering that suggests that the flagellar systems and Type III protein secretory systems diverged from each other following very early duplication of a gene cluster sharing many (but not all) genes. Phylogenies of most or all of the flagellar proteins follow those of the source organisms with little or no lateral gene transfer suggesting that homologous flagellar proteins are true orthologues. We suggest that the flagellar apparatus was the evolutionary precursor of Type III protein secretion systems.

Introduction

Bacteria use a plethora of mechanisms for the secretion of macromolecules from the site of synthesis (the cell cytoplasm) to their sites of action. Comprehensive analyses of these systems have revealed the occurrence of seven types of systems that allow transfer of intact proteins across the cytoplasmic membranes of Gram-negative bacteria (the ABC, Sec, Vir, Conj, Tat, MscL and holin-type systems; Table 1) and nine types of systems that facilitate transport of these molecules across the outer membranes of these organisms (the Vir, Conj, MTB, FUP, AT, OMR, OMF, TEC and secretin-type systems; Table 1) (Saier, 1998; Saier, 1999) (see our website [http://www-biology.ucsd.edu/~msaier/transport/]). While some of these systems are found ubiquitously within the three domains of living organisms, the majority appears to be bacterial or even Gram-negative bacterial specific. Some of these systems are simple, consisting of just one or a few proteins, but others are complex, consisting of over a dozen constituents. While the former types are often protein specific, the latter types generally transport multiple protein species (Anderson *et al.,* 1999; Rossier *et al.,* 1999; Young *et al.,* 1999).

Perhaps the most complex of these are the Type III protein secretion systems (TIIIPS; Vir type; TC #3.A.6) which function in the nearly exclusive export of virulence factors in Gram-negative bacteria (Lee, 1997; Hueck, 1998; Galán and Collmer, 1999). These complex systems can catalyze the translocation of proteins into the extracellular medium, the cytoplasmic membrane of the host cell, or directly into the host cell cytoplasm (Hueck, 1998; Galán and Collmer, 1999; Lee and Schneewind, 1999; Wachter *et al.,* 1999). They thus can transfer proteins across two or three membranes: the two membranes of the Gram-negative bacterial cell envelope and the cytoplasmic membrane of the host animal or plant cell. Although it is frequently assumed that TIIIPS systems function via a one-step secretion mechanism involving a single energy-coupled step, there is currently little evidence to support this postulate.

The organization of Type III protein secretory system-encoding genetic apparatuses is shown in Figure 1. All TIIIPS constituents of a particular system are usually (but not always) encoded within a single gene cluster. Moreover, in many of these clusters, similar gene orders are observed (for a detailed discussion of these similarities see Hueck, 1998). These TIIIPS system genes are often found together with those encoding secreted or potentially secreted proteins, chaperone proteins specific for these secreted proteins and transcriptional regulatory proteins controlling expression of the represented operons. Only the TIIIPS system genes in *Chlamydia* and the flagellar biosynthetic genes in *E. coli* are not located contiguously (Figure 1).

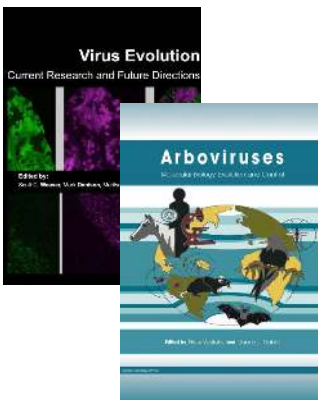
TIIIPS systems have been shown to exhibit broad specificity with respect to their protein substrates (Hermant *et al.,* 1995; Rosqvist *et al.,* 1995; Frithz-Lindsten *et al.,* 1997; Anderson *et al.,* 1999; Rossier *et al.,* 1999). Two possibilities exist with respect to the mechanism of transfer: (1) the proteins of a Type III secretory system could function in a coordinated, cooperative process in which all constituents are interconnected due to dependencies on extensive protein-protein interactions, or (2) the proteins might function as a set of subcomplexes that catalyze the

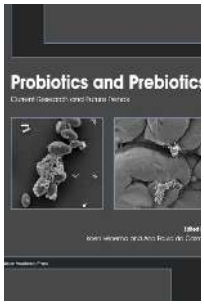
Received January 31, 2000; accepted January 31, 2000. *For correspondence. Email msaier@ucsd.edu; Tel. (858) 534-4084; Fax. (858) 534-7108.

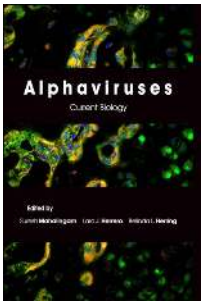
Further Reading

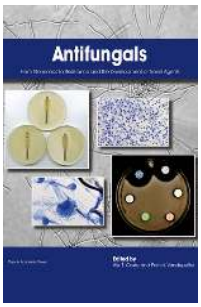
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem
- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo
- ***Shigella*: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile









- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant
- ***Aeromonas***
  Edited by: J Graf (2015)
  www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics
- ***Leishmania*: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2
- ***Acanthamoeba*: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order

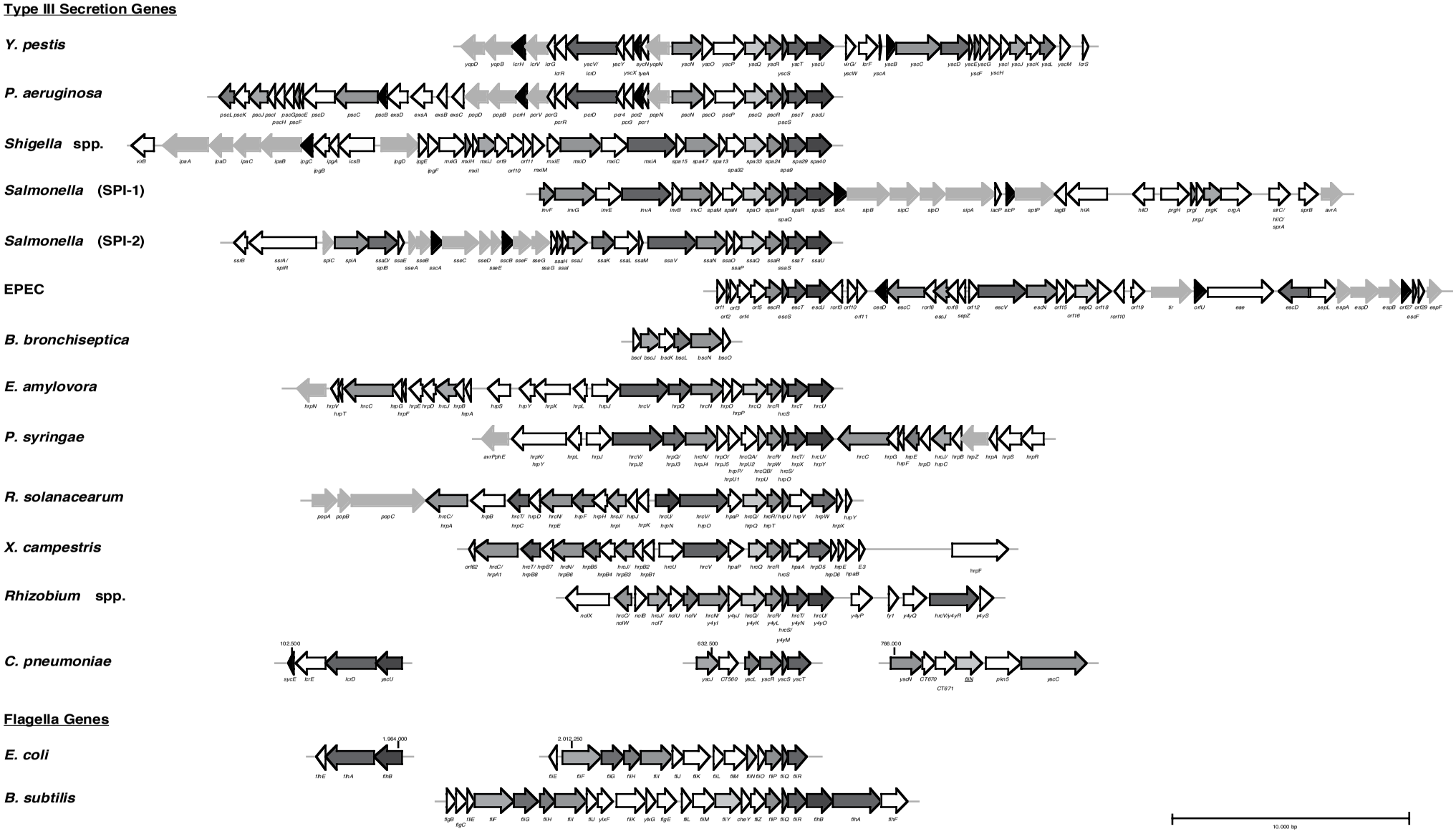


Figure 1. Schematic overview of the genes encoding Type III protein secretion systems in various bacterial pathogens and homologous flagellar biosynthetic genes in *E. coli* and *B. subtilis*. Arrows denote genes and their direction of transcription. The name of each gene is shown underneath the arrow. In cases of different names for the same gene, these names are shown in columns, separated by /. The more actual name is shown above. A scale (in bp) is shown. Hatched arrows: Genes encoding broadly conserved proteins of the secretion apparatus are marked with the same hatching. Filled grey arrows: Genes that encode secreted or potentially secreted proteins. Filled black arrows: Genes encoding chaperones or potential chaperones for secreted proteins. The Type III secretion genes of *Chlamydia* spp. and the flagellar biosynthetic genes of *E. coli* are not located on contiguous stretches of DNA. For these genes, small numbers above the genes denote the positions of these genes in bp within the completely sequenced genomes of these organisms. In *Chlamydia*, the flagellar biosynthetic gene *fliN* is located in proximity to the *yscC* gene.

Table 1. Protein Translocases in Gram-Negative Bacteria and Eukaryotic Organelles

PT#	Type	Name	TC#	Bacteria	Archaea	Eukarya	#Proteins	Energy
I	ABC	ATP-binding cassette translocase	3.A.1	+	+	+	1-2	ATP
II	Sec	General secretory translocase	3.A.5	+	+	+	~12	GTP or ATP+pmf
III	Vir	Virulence-related translocase	3.A.6	+	-	-	>10	ATP
IV	Conj	Conjugation-related translocase	3.A.7	+	-	-	>6	ATP
V	Tat	Twin arginine targeting translocase	2.A.64	+	+	+ (chloroplasts)	5	pmf
VI	MPT	Mitochondrial protein translocase	3.A.8	-	-	+ (mitochondria)	~20	ATP
VII	CEPT	Chloroplast envelope protein translocase	3.A.9	+	-	+ (chloroplasts)	3	GTP
VIII	MTB	Main terminal branch of the Sec translocase (Gram-negative bacterial outer membranes)	3.A.5	+	-	-	~14	ATP?
IX	Channels	Cytoplasmic membrane channels						
	(a) Bcl-2	Bcl-2 family	1.A.21	-	-	+		None
	(b) MscL	Large conductance mechano-sensitive channel family	1.A.22	+	-	-		None
	(c) Holins	Holin functional superfamily	1.A.28-1.A.43	+	-	-		None
X	OMPs	Gram-negative bacterial outer membrane channel-forming translocases						
	(a) FUP	Fimbrial usher protein	1.B.11	+	-	-	1	None
	(b) AT	Auto transporter	1.B.12	+	-	-	1	None
	(c) OMR	Outer membrane receptor	1.B.14	+	-	-	1	None
	(d) OMF	Outer membrane factor	1.B.17	+	-	-	1	None
	(e) TEC	Toxin export channel	1.B.20	+	-	-	1	None
	(f) Secretins	Secretin	1.B.22	+	-	-	1	None
XI	Toxins	Channels targeted to cells other than ones that synthesize them						
	(a) DT[a]	Diphtheria toxin family	1.C.7	+	-	-		None
	(b) BTT[a]	Botulinum and tetanus toxin family	1.C.8	+	-	-		None
	(c) IIITCP[a]	Bacterial type III-target cell pore family	1.C.36	+	-	-		None

[a]Made by bacteria but inserted into host animal or plant cell membranes. **http://www-biology.ucsd.edu/~msaier/transport/titlepage.html**

overall secretory process as a series of independent steps, each with relative autonomy. If the latter mechanism were operative, then each subcomplex could undergo evolutionary divergence independently of the other subcomplexes. However, if the former mechanism were operative, the entire system would be expected to undergo functional divergence in a mutualistic process that would prevent homologous sequence divergent proteins from substituting for each other without loss of transport efficiency. Such protein-protein interdependencies would thus prevent exchange of constituents between systems once a certain degree of sequence divergence had been attained.

In order to attempt to distinguish these possibilities, we have conducted phylogenetic analyses of all essential constituents of Type III systems and their homologues. These homologues include basal body constituents of the bacterial flagellum that have been reported to exhibit the capacity to transport virulence factors (Young *et al.,* 1999). The results of our phylogenetic analyses show that little or no exchange of constituents between systems has occurred, even when two or more systems are present in a single bacterial cell. The results of these analyses support a cooperative, concerted mechanism of secretion.

Proteins of Type III Protein Secretion (TIIIPS) Systems and Their Homologues

Table 2 summarizes the available structural, functional and subcellular localization information about components of the *Yersinia* virulence-related, Type III secretion system which are broadly conserved in other TIIIPS systems and the homologous flagellar proteins in *S. typhimurium* and *E. coli* (see also Figure 1 and Hueck, 1998 for further discussion). The broadly conserved components of the *Yersinia* Ysc system comprise 12 proteins varying in size between 87 residues and 704 residues. Six of these proteins are predicted to be in the inner membrane; three may be cytoplasmic; one is a lipoprotein, possibly connecting the inner and outer membranes; and one is the outer membrane secretin that forms the transmembrane pore in this structure. The twelve non-homologous Ysc protein constituents which are broadly conserved in other TIIIPS systems were included in the phylogenetic analyses to be reported here (Tables 2 and 3). Nine of these proteins have homologues in the flagellar complex, and one protein family, the secretins, lacking in the flagellar complex, include numerous homologues that are involved in Types II (TIIPS) and III protein secretion, in fimbrial (pilin) export, and in phage assembly and export (Table 3). A few secretins are involved in still other processes such as competence for DNA uptake, export of toxins, and nodulation of leguminous plants by N_2-fixing bacteria (Table 3). Table 3 also shows the unified nomenclature for the broadly conserved TIIIPS constituents, with the general designation Sct for secretion and cellular translocation, as has been proposed earlier (Hueck, 1998).

Table 4 lists the organisms alphabetically and indicates the type of pathogen (plant, animal or bacterial) as well as the type of secretion system present (Types II (G) and III (T) secretion systems; pilin export (P); flagellar protein export (F); bacteriophage assembly (V); nodulation (N), and competence (C)). It is noteworthy that several organisms have more than one system, and that four systems have been identified in both *E. coli* and *P. aeruginosa* (general (G) and Type III (T) secretion systems as well as flagellar (F) and fimbrial (pilin; P) systems).

Table 2. Broadly Conserved Components of the *Yersinia* Type III Protein Secretion System and Homologues in the Bacterial Flagellar System

Conserved *Yersinia* or TIIIPS protein[a]	LcrD	YscN	YscQ	YscR	YscS	YscT	YscU	YscC	YscD	YscF	YscJ	YscL
Size of *Yersinia* TIIIPS Protein[b]	704	439	307	217	88	261	354	607	418	87	244	223
Size range of homologues[b]	582-733	430-473	102-382	172-306	76-95	251-289	345-376	412-921	310-432	71-87	190-599	193-314
Fla protein in Eco or Sty[c]	FlhA (Ex App)	FliI (ATPase)	FliNM (C-ring)	FliP (Ex App)	FliQ (Ex App)	FliR (Ex App)	FlhB (Ex App)	-	-	-	FliF (MSring)	FliH (Ex App)
Size of Fla protein in Eco or Sty[b,c]	694 (Sty)	456 (Sty)	N, 137 M, 344 (Eco)	245 (Eco)	89 (Eco)	261 (Eco)	383 (Sty)				560 (Sty)	235 (Sty)
Probable location[d]	IM	Cyt	?	IM	IM	IM	IM	OM		Cyt	LP	Cyt?
# putative TMSs[e]	8	0	0-2?	4	2	6	4	Secretin (ß)	1-2	0	1-3	0
Proposed function[d]	IM pore?		Energize secretion or assembly				Regulation	OM pore			Bridge btw IM and OM	

[a]All proteins except LcrD are designated Ysc. Only those *Yersinia* TIIIPS proteins are included which have sequenced hornologues in several other TIIIPS systems.
[b]Protein sizes are indicated in numbers of amino acids.
[c]Eco *E. coli;* Sty, *S. typhimurium;* Ex App, Export Apparatus; C-ring, the flagellar basal body ring exposed to the cytoplasm; MS-ring the flagellar basal body ring embedded in the cytoplasmic membrane.
[d]The abbreviations are: Cyt. cytoplasm, IM, inner memhrane; OM, outer membrane; LP. lipoproteins which may bridge inner and outer membranes.
[e]# putative TMSs, number of putative α-helical transmembrane spanners; ß, ß-structure. YscR has its N- and C-termini in the periplasm; YscS and YscT have their N- and C-termini in the cytoplasm (see Hueck, 1998).

16S Ribosomal RNA Phylogenetic Trees

In order to provide a reference point for organismal phylogeny, two phylogenetic trees were constructed based on 16S ribosomal RNAs from a representative species of each bacterial genus that (1) possesses a Type III protein secretion system (Figure 2; see Table 3 for genuses included and organismal abbreviations) or possesses flagellar proteins included in this study (Figure 3; see Table 3). The genuses represented in these figures and the organismal 3-letter abbreviations used in these figures are presented in Table 3. At the bottom of the tree shown in Figure 2 are close relatives of the enteric bacteria: *Shigella flexneri*, *Escherichia coli*, *Salmonella typhimurium*, *Erwinia amylovora* and *Yersinia enterocolitica*. The *Pseudomonas aeruginosa* 16S rRNA clusters very loosely with those from the enteric bacteria. Also as expected, *Rhizobium spp.* clusters tightly with *Sinorhizobium fredii* while *Burkholderia pseudomallei*, *Ralstonia solanacearum* and *Bordetella bronchiseptica* cluster loosely together. The other organisms represented (*Chlamydia trachomatis* and *Xanthomonas campestris*) do not exhibit significant clustering as these two branches emanate from points near the center of the unrooted tree. These relationships are taken as an indication of organismal phylogeny (Olsen *et al.,* 1994).

Figure 3 presents a comparable 16S rRNA tree for the genuses from which sequenced flagellar proteins were obtained. Two major clusters and two minor clusters can be observed. The first major cluster (bottom of Figure 3) includes the rRNAs of the enteric bacteria and their close relatives including *S. typhimurium* and *E. coli*, *Y. enterocolitica* and *Erwinia carotovara*, *Proteus mirabilis*, *Vibrio parahaemolyticus*, *Legionella pneumophila*, and *Pseudomonas aeruginosa* with increasing distance from *S. typhimurium* in that order. The second major cluster (top of Figure 3) includes *Sinorhizobium meliloti* and *Agrobacterium tumefaciens* clustering tightly together with increasing distances from *Brucella abortus*, *Rhodobacter sphaeroides*, *Caulobacter crescentus* and *Zymomonas mobilis* in this order. The two small clusters, each with only two members, include the two spirochetes, *Treponema pallidum* and *Borrelia burgdorferi* (left hand side of Figure 3), and *Campylobacter jejuni* and *Helicobacter pylori* (right hand side of Figure 3). The remaining organisms, *Bacillus subtilis*, *Aquifex aeolicus* and *Chlamydia trachomatis,* are localized to branches that stem from points near the center of the tree. These relationships are assumed to reflect organismal phylogenies (Olsen *et al.,* 1994).

Protein Phylogenetic Trees

Phylogenetic trees were constructed using two different programs, and one of the programs was applied to two different subsets of proteins. (1) The TREE program (Feng and Doolittle, 1990) was used to construct trees for the LcrD, YscC, YscD, YscN, and YscQ-U families. (2) The ClustalX program (Thompson *et al.,* 1997) was used to construct trees for (a) those protein families for which the largest numbers of homologues are available in both the TIIIPS and flagellar (Fla) systems (LcrD, YscC, YscJ, YscN, and YscQ-U families). In this first set of trees we included only proteins from those TIIIPS and Fla systems for which the sequences of all or all but one member of these systems are available (indicated in Table 3 in bold print). The ClustalX program was also used to (b) construct extended trees for all protein families listed in Table 3, including essentially all fully sequenced protein homologues for each of these twelve families. Only partially sequenced proteins and homologues which were extremely similar in sequence (>90% identical) to one that was included were omitted from this last study. The omitted proteins are indicated by the superscript letter “c” in Tables 5-16 (see footnotes to Table 5) and by asterisks in Table 3. Trees that were

Table 3. Proteins Included in This Study[a]

Yersinia nomenclature:	LcrD	YscN	YscQ	YscR	YscS	YscT	YscU	YscC	YscD	YscF	YscJ	YscL
Unified nomenclature for TIIIPS system constituents:	SctV	SctN	SctQ	SctR	SctS	SctT	SctU	SctC	SctD	SciF	SctJ	SctL
Abbr.	Organism											
	Type III protein secretion											
Bps	*Burkholderia pseudomallei*	?	?	SctQ	SctR	SctS	?	?	?	?	?	?	?
Bbr	*Bordetella bronchiseptica*	?	BscN	?	?	?	?	?	?	?	?	BscJ	BscL
Cca	*Chlamydia caviae*	Cds2*	?	?	?	?	?	Cds2*	?	?	?	?	?
Cpn	*Chlamydia pneumoniae*	**LcrD***	YScN*	-	**YscR***	**YscS***	**YscT***	**YscU***	**YscC***	-	-	YScJ*	YScL*
Ctr	*Chiamydia trachomatis*	**LcrD**	YscN	YscQ	**YscR**	**YscS**	**YscT**	**YscU**	**YscC**	-	-	YscJ	YscL
Eam	*Erwinia amvlovora*	**HrpI**	HrcN	HrcQ	**HrcR**	**HrcS**	**HrcT**	**HrcU**	**HrcC**	**HrpQ**	?	HrcJ	HrpE
Ech	*Erwinia chrysanthemi*	?	?	?	?	?	?	?	HrcC*	?	?	?	?
Ehe	*Erwinia herbicola*	?	?	HrcQ2*	HrcR*	HrcS*	HrcT*	HrcU*	?	?	?	?	?
Eco	*Escherichia coli*	**EscV**	EscN	-	**EscR**	**EscS**	**EscT**	**EscU**	**EscC**	EscD	EscF	EscJ	-
Eco	*Escherichia coli*	-	-	-	-	InvX	-	-	-	-	-	-	-
Pae	*Pseudomonas aeruginosa*	PcrD	PscN	?	?	?	?	?	PscC	PscD	PscF	PscJ	PscL
Psy	*Pseudomonas syringae (gly)*	?	?	?	?	?	?	HrcU*	?	?	?	HrcJ	HrpE
Psy	*Pseudomonas syringae (phs)*	?	?	HrcQ2	HrcR	HrcS	HrcT	HrcU	?	?	?	?	?
Psy	*Pseudomonas syringae (syr)*	HrpI	HrpJ4	HrpU*	?	?	?	HrpY*	HrcC	?	?	HrpC*	-
Rso	*Ralstonia solanacearum*	**HrpQ**	HrpE	HrpQ	**HrpT**	**HrpU**	**HrpC**	**HrpN**	**HrpA**	-	-	HrpI	HrpF
Rhi	*Rhizobium spp*	**Y4yR**	Y4yI	Y4yK	**Y4yL**	**Y4yM**	**Y4yN**	**Y4yO**	**Y4yJ**	-	-	?	?
Sdu	*Salmonella dublin*	?	?	SpaO*	?	SpaQ*	?	?	?	?	?	?	?
Sen	*Salmonella enterica*	InvA*	?	SpaO	SpaP	SpaQ	?	?	?	?	?		
Sti	*Salmonella typhi*	?	?	SpaO*	SpaP*	?	?	?	?	?	?	?	?
Sty	*Salmonella typhimurium SPII*	**SsaV**	SsaN	SsaQ	**YscR**	**SsaS**	**SsaT**	**SsaU**	**SpiA**	SpiB	SsaH	SsaJ	SsaK
Sty	*Salmonella typhimurium SPI2*	**InvA**	SpaL	SpaO	**SpaP**	**SpaQ**	**SpaR**	**SpaS**	**InvG**	SsaD	PrgI	PrgK	-
Sfl	*Shigella flexneri*	**MxiA**	SpaL	?	**SpaP**	**SpaQ**	**SpaR**	**SpaS**	**MxiD**	-	MxiH	MxiJ	-
Sso	*Shigella sonnei*	Orf*	?	?	Orf*	?	?	MxiD*	?	?	MxiJ*	?	
Sfr	*Sinorhizobium fredii*	?	HrcN*	HrcQ*	HrcR*	HrcS*	HrcT*	HrcU*	?	?	?	?	?
Xca	*Xanthomonas campestris*	**HrcP2**	HrpB6	HrcQ	**HrcR**	**HrcS**	**HrpB8**	?	**HrpAI**	HrpD5	HrpB3HrpB5		
Xor	*Xanthomonas oryzae*	?	?	?	?	?	?	?	**HrpA***	?	?	?	?
Yen	*Yersinia enterocolitica*	**LcrD**	YscN	YscQ	YscR	YscS	YscT	YscU	YscC	YscD	YscF	YscJ	YscL
Yen	*Yersinia enterocolitica*	YsaA	-	-	-	-	-	-	-	-	-	-	-
Ype	*Yersinia pestis*	**LcrD***	YscN*	YscQ*	**YscR***	**YscS***	**YscT***	**YscU***	**YscC***	YscD*	YscF*	YScJ*	YScL*
Yps	*Yersinia pseudotuberculosis*	**LcrD***	YScN*	YScQ*	**YscR***	**YscS***	**YscT***	**YscU***	?	?	?	YScJ*	YScL*
	Flagellar												
Atu	*Agrobacterium tumefaciens*	?	FliI	FliN	FliP	?	?	FlhB	-	-	-	?	?
Aae	*Aquifex aeolicus*	**FlhA**	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliF	-
Bsu	*Bacillus subtilis*	**FlhA**	FliI	FliY	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliF	FliH
Bbu	*Borrelia burgdorferi*	**FlhA**	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliF	FliH
Bab	*Brucella abortus*	?	?	?	?	?	?	?	-	-	-	FliF	?
Cje	*Campylobacter jejuni*	FlhA	?	FliN	?	?	?	FlhB	-	-	-	?	?
Ccr	*Caulobacter crescentus*	**FlhA**	FliI	FliN	**FliP**	**FliQ**	**FliR**	**PodW**	-	-	-	FliF	?
Cpn	*Chlamydia pneumoniae*	FlhA*	-	-									
Ctr	*Chlamydia trachomatis*	FlhA	-										
Eca	*Erwinia carotovara*	?	?	FliN	FliP	FliQ	FliR	?	-	-	-	?	?
Eco	*Escherichia coli*	**FlhA**	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliL	FliH
Eco	*Escherichia coli*	FhiA	-	-	-	-	-	-	-	-	-	-	-
Hpy	*Helicobacter pylori*	**FlhA**	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliF	FliH
Lpn	*Legionella pneumophila*	?	FliI	?	?	?	?	?	-	-	-	?	?
Pmi	*Proteus mirabilis*	FlhA	?	?	?	?	?	FlhB*	-	-	-	?	?
Pae	*Pseudomonas aeruginosa*	?	?	FliN	FliP	?	?	?	-	-	-	FliF	?
Ppu	*Pseudomonas putida*	FlhA*	?	?	FliP*	?	FliR*	?	-	-	-	?	?
Rsp	*Rhodobacter sphaeroides*	?	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliF	?
Sty	*Salmonella typhimurium*	**FlhA**	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	FliF	FliH
Sme	*Sinorhizobium meliloti*	?	FliI	FliN	FliP	?	?	FlhB	-	-	-	FliF	?
Tde	*Treponema denticola*	-	FliI*	?	FliP*	-	-	-	-	-	-	FliF*	FliH*
Tpa	*Treponema pallidum*	FlhA	FliI	?	FliP	FliQ	FliR	FlhB	-	-	-	FliF	FliH
Vpa	*Vibrio parahaemolyticus*	FlhA	FliI	FliN	**FliP**	**FliQ**	**FliR**	**FlhB**	-	-	-	-	FliH
Yen	*Yersinia enterocolitica*	FlhA	?	?	?	?	?	FlhB	-	-	-	?	?
Zmo	*Zymomonas mobilis*	FlhA	?	?	?	?	?	?	-	-	-	FliF	?
	Type II protein secretion												
Ahy	*Aeromonas hydrophila*	-	-	-	-	-	-	-	GspD	-	-	-	-
Ahy	*Aeromonas hydrophila*	-	-	-	-	-	-	-	SpsD	-	-	-	-
Asa	*Aeromonas salmoncida*	-	-	-	-	-	-	-	GspD*	-	-	-	-
Aae	*Aquifex aeolicus*	-	-	-	-	-	-	-	GspD	-	-	-	-
Bps	*Burkholderia pseudomallei*	-	-	-	-	-	-	-	GspD	-	-	-	-
Cpn	*Chlamydia pneumoniae*	-	-	-	-	-	-	-	GspD*	-	-	-	-
Ctr	*Chlamydia trachomatis*	-	-	-	-	-	-	-	GspD	-	-	-	-
Cli	*Chlorobium limicola*	-	-	-	-	-	-	-	Exp	-	-	-	-
Eca	*Erwinia carotovora*	-	-	-	-	-	-	-	GspD	-	-	-	-
Ech	*Erwinia chrysanthemi*	-	-	-	-	-	-	-	GspD*	-	-	-	-
Eco	*Escherichia coli*	-	-	-	-	-	-	-	EtpD	-	-	-	-
Eco	*Escherichia coli*	-	-	-	-	-	-	-	GspD	-	-	-	-
Eco	*Escherichia coli*	-	-	-	-	-	-	-	HofQ*		-	-	-
Kpn	*Klebsiella pneumoniae*	-	-	-	-	-	-	-	GspD	-	-	-	-
Pae	*Pseudomonas aeruginosa*	-	-	-	-	-	-	-	GspD	-	-	-	-
Pae	*Pseudomonas aeruginosa*	-	-	-	-	-	-	-	XqhA	-	-	-	-
Syn	*Synechocystis spp.*	-	-	-	-	-	-	-	GspD	-	-	-	-
Vch	*Vibrio cholerae*	-	-	-	-	-	-	-	GspD	-	-	-	-
Xca	*Xanthomonas campestris*	-	-	-	-	-	-	-	GspD	-	-	-	-
	Pilin secretion												
Aae	*Aquifex aeolicus*	-	-	-	-	-	-	-	Orf	-	-	-	-
Cbu	*Coxiella bunietii*	-	-	-	-	-	-	-	Orf*		-	-	-
Eco	*Escherichia coli*	-	-	-	-	-	-	-	BfpB	-	-	-	-
Mxa	*Myxococcus xanthus*	-	-	-	-	-	-	-	PilQ	-	-	-	-
Ngo	*Neisseria gonorrhoeae*	-	-	-	-	-	-	-	PilQ	-	-	-	-
Nme	*Neisseria meningitidis*	-	-	-	-	-	-	-	PilQ*	-	-	-	-
Pac	*Pseudomonas aeruginosa*	-	-	-	-	-	-	-	PilQ	-	-	-	-
Pal	*Pseudomonas alcaligenes*	-	-	-	-	-	-	-	XcpQ*	-	-	-	-
Ppu	*Pseudomonas putida*	-	-	-	-	-	-	-	XcpQ*	-	-	-	-
Vch	*Vibrio cholerae*	-	-	-	-	-	-	-	TcpC	-	-	-	-
	Bacteriophage assembly												
BPfl	*Bacteriophage fi*	-	-	-	-	-	-	-	VG4	-	-	-	-
BPfd	*Racteriophage fd*	-	-	-	-	-	-	-	VG4	-	-	-	-
BPfs2	*Bacteriophage fs-2*	-	-	-	-	-	-	-	VG4	-	-	-	-
BP122	*Bacteriophage I2-2*	-	-	-	-	-	-	-	VG4	-	-	-	-
BPIf1	*Bacteriophage Ifi*	-	-	-	-	-	-	-	VG4	-	-	-	-
BPIke	*Bacteriophage Ike*	-	-	-	-	-	-	-	VG4	-	-	-	-
BPM13	*Bacteriophage M13*	-	-	-	-	-	-	-	VG4	-	-	-	-
BPPf3	*Bacteriophage Pf3*	-	-	-	-	-	-	-	VG4	-	-	-	-
	Nodulation												
Rhi	*Rhizobium spp.*	-	-	-	-	-	-	-	NolW	-	-	NolT	No1V
Sfr	*Sinorhizobium fredii*	-	-	-	-	-	-	-	NolW*	-	NolT	-	
	Competence												
Hin	*Haemophilus influenzae*	-	-	-	-	-	-	-	ComE	-	-	-	-

[a]Proteins in bold were included in the phylogenetic trees of Figure 1. Proteins indicated with an asterisk (*) were excluded from the phylogenetic trees of Figures 2-13.

Table 4. Species Analyzed for Type III Protein Secretion Systems and Homologues

Organism	Abbreviation	Type of pathogen[a]	Type of system(s) [b]
Bacteria			
Aeromonas hydrophila	Ahy	Animal	G
Aeromonas salmoncida	Asa	Animal	G
Aquifex aeolicus	Aae	-	F, G, P
Bacillus subtilis	Bsu	-	F
Bordetella bronchiseptica	Bbr	Animal	T
Brucella abortus	Bab	Animal	F
Burkholderia pseudomallei	Bps	Plant	T, G
Campylobacter jejuni	Cje	Animal	F
Caulobacter crescentus	Ccr	-	F
Chlamydia caviae	Cca	Animal	T
Chlamydia pneumoniae	Cpn	Animal	T, F, G
Chlamydia trachomatis	Ctr	Animal	T, F, G
Chlorobium limicola	Cli	-	G
Coxiella burnetti	Cbu	Animal	P
Erwinia amylovora	Eam	Plant	T
Erwinia carotovora	Eca	Plant	F, G
Erwinia chrysanthemi	Ech	Plant	T, G
Erwinia herbicola	Ehe	Plant	T
Escherichia coli	Eco	Animal	T, F, G, P
Haemophilus influenzae	Hin	Animal	C
Helicobacter pylori	Hpy	Animal	F
Klebsiella pneumoniae	Kpn	Animal	G
Legionella pneumophila	Lpn	Animal	F
Myxococcus xanthus	Mxa	Plant?	P
Neisseria gonorrhoeae	Ngo	Animal	P
Neisseria meningitidis	Nme	Animal	P
Proteus mirabilis	Pmi	Animal	F
Pseudomonas acaligenes	Pac	Animal?	G
Pseudomonas aeruginosa	Pae	Animal	T, F, G, P
Pseudomonas putida	Ppu	?	F, G
Pseudomonas syringae	Psy	Plant	T
Ralstonia solanacearum	Rso	Plant	T
Rhizobium spp.	Rhi	Plant	T, N
Rhodobacter sphaeroides	Rsp	-	F
Salmonella dublin	Sdu	Animal	T
Salmonella enterica	Sen	Animal	T
Salmonella typhi	Sti	Animal	T
Salmonella typhimurium	Sty	Animal	T, F
Shigella flexneri	Sfl	Animal	T
Shigella sonnei	Sso	Animal	T
Sinorhizobium fredii	Sfr	Plant	T, N
Sinorhizobium meliloti	Sme	Plant	F
Synechocystis spp.	Syn	-	G
Treponema denticola	Tde	Animal	F
Treponema pallidum	Tpa	Animal	F
Vibrio cholerae	Vch	Animal	G, P
Vibrio parahaemolyticus	Vpa	Animal	F
Xanthomonas campestris	Xca	Plant	T, G
Xanthomonas oryzae	Xor	Plant	T
Yersinia enterocolitica	Yen	Animal	T, F
Yersinia pestis	Ype	Animal	T
Yersinia pseudotuberculosis	Yps	Animal	T
Zymomonas mobilis	Zmo	Plant?	F
Viruses			
Bacteriophage f1	BPf1	Bacteria	V
Bacteriophage fd	BPfd	Bacteria	V
Bacteriophage fs-2	BPfs2	Bacteria	V
Bacteriophage I2-2	BPI22	Bacteria	V
Bacteriophage If1	BPIf1	Bacteria	V
Bacteriophage Ike	BPIke	Bacteria	V
Bacteriophage M13	BPM13	Bacteria	V
Bacteriophage Pf3	BPPf3	Bacteria	V

[a]The bacteria are pathogens or symbionts of plants or animals, as indicated above, or they are not pathogens (-). Phage are considered as bacterial pathogens.

[b]Systems are designated as follows: Type III protein secretion, T; Flagellar, F; General (Type II protein secretion), G; Pilin secretion, P; Phage assembly and export, V; Nodulation, N; Competence, C.

Type III PS 16S rRNA tree

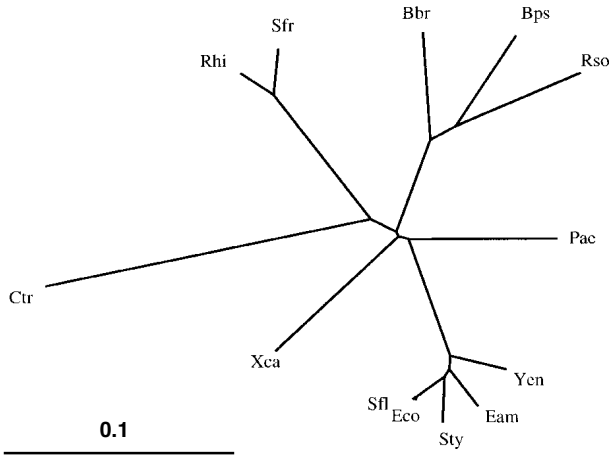


Figure 2. Phylogenetic tree of 16S rRNAs derived from Gram-negative bacterial genuses that include Type III protein secretion systems tabulated in Table 3 and included in this study.

constructed with both the TREE and ClustalX programs (LcrD, YscC, and YscR-U families) generally yielded good agreement. In both sets, the Fla and TIIIPS proteins clustered separately, and clustering within each of these two groups of proteins was similar. Consequently, only one of these sets of trees, that produced with the ClustalX program, will be presented.

Flagellar 16S rRNA tree

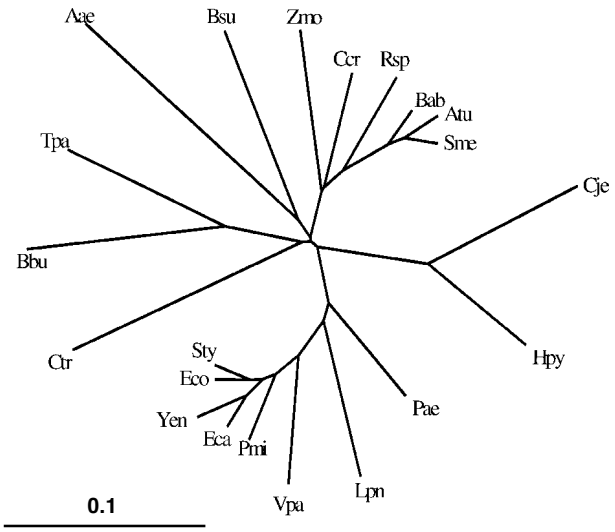


Figure 3. Phylogenetic tree of 16S rRNAs derived from the bacterial genuses that include flagellar proteins tabulated in Table 3 and included in this study.

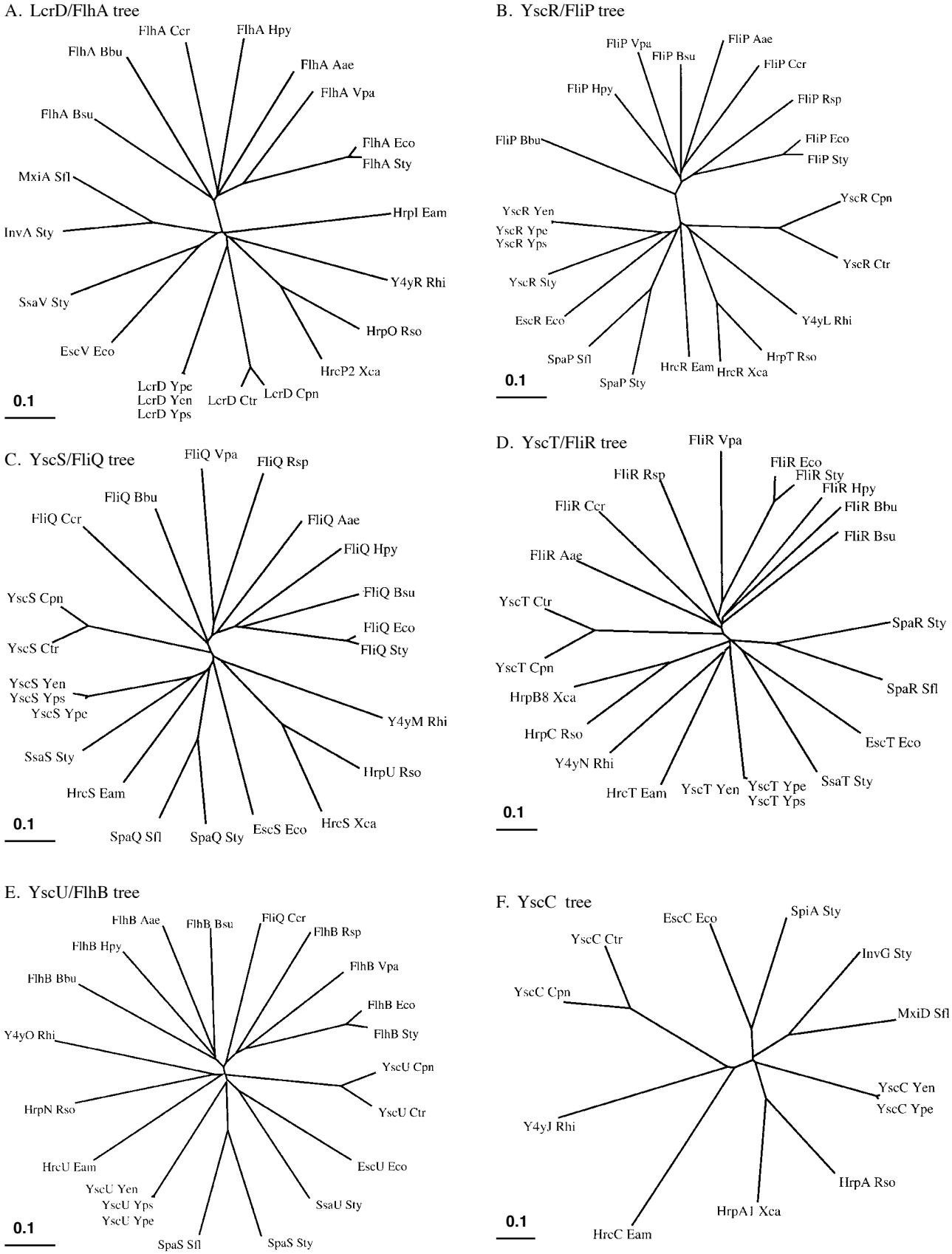


Figure 4. Phylogenetic trees for six families of TIIIPS-Fla homologues. Protein constituents from only the most completely sequenced TIIIPS systems are represented. Thus, the following proteins are missing from the trees shown in Figures A-F as indicated in Table 3. A: The *R.spheroides* FlhA homologue; B-D, no proteins are missing; E, the *X. campestris* Hrp/Hrc homologue is missing; F: all secretin homologues are lacking. The ClustalX program (Thompson *et al.*, 1997) was used to generate these trees as well as those portrayed in Figures 5-16.

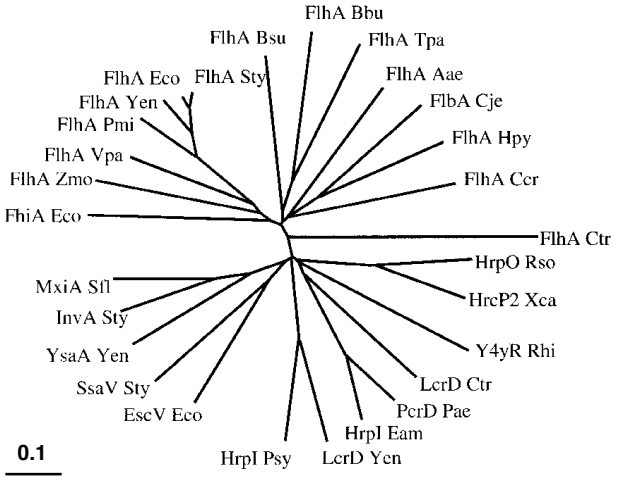


Figure 5. Phylogenetic tree for the LcrD/FlhA family.

Phylogenetic Trees for Homologous Constituents of Completely Sequenced TIIIPS and Fla Systems

For the ClustalX trees that were constructed for the LcrD, YscC, YscJ, YscN, and YscQ-U families from the complete or nearly complete TIIIPS and Fla systems analyzed, clustering was comparable for all nine trees. Six of these trees (LcrD, YscC, and YscR-U families) are presented in Figures 4A-F. The specific observations derived from these trees are summarized as follows: (1) In all nine trees, the Fla proteins (when present) clustered separately from the TIIIPS proteins, except that in the YscQ tree (not shown), FliN Aae clustered loosely with the TIIIPS proteins. No Fla proteins were included in the YscC tree as the flagellar apparatus lacks secretin homologues (Figure 1F; Table 3). (2) The two chlamydial TIIIPS proteins (*C. pneumoniae* and *C. trachomatis*) always clustered tightly together, as did the three TIIIPS *Yersinia* proteins and the two Fla proteins of *E. coli* and *S. typhimurium*. (3) The TIIIPS *R. solanacearum* and *X. campestris* proteins as well as the Inv/Spa proteins of the first TIIIPS system of *S. typhimurium*

Table 5. LcrD/FlhA Group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FlhA Aae	Flagellar export protein	*Aquifex aeolicus*	678	2983661	gbAE000729
FlhA Bsu	Flagellar export protein	*Bacillus subtilis*	677	544312	spP35620
FlhA Bbu	Flagellar export protein	*Borrelia burgdorferi*	673	1165256	gbU43739
FlbA Cje	Flagellar biosynthesis protein	*Campylobacter jejuni*	724	477572	pirA49217
FlhA Ccr	Flagellar export protein	*Caulobacter crescentus*	700	462103	spQ03845
Cds2 Cca[c]	Type III secretion protein	*Chlamydia caviae*	709	2358257	gbU88070
LcrD Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	710	4376601	gbAE001617
FlhA Cpn[c]	Flagellar biosynthesis protein	*Chlamydia pneumoniae*	582	4376639	gbAE001620
LcrD Ctr	Type III secretion protein	*Chlamydia trachomatis*	708	3328486	gbAE001283
FlhA Ctr	Flagellar biosynthesis protein	*Chlamydia trachomatis*	605	3328453	gbAE001280
HrpI Eam	Type III secretion protein	*Erwinia amylovora*	697	547673	spP35654
EscV Eco	Type III secretion protein	*Escherichia coli*	675	2865289	gbAF022236
FhiA Eco	Inner membrane transporter	*Escherichia coli*	579	2494467	spQ47153
FlhA Eco	Flagellar biosynthesis protein	*Escherichia coli*	692	2494468	spP76298
FlhA Hpy	Flagellar biosynthesis protein	*Helicobacter pylori*	733	2494469	spO06758
FlhA Pmi	Flagellar biosynthesis protein	*Proteus mirabilis*	696	2494470	spQ51910
PcrD Pae	Type III secretion protein	*Pseudomonas aeruginosa*	706	2459978	gbAF010150
FlhA Ppu[c]	Flagellar biosynthesis protein	*Pseudomonas putida*	223	2853594	gbAF031898
HrpI Psy	Type III secretion protein	*Pseudomonas syringae (syr)*	695	547674	spP35655
HrpO Rso	Type III secretion protein	*Ralstonia solanacearum*	690	547677	spP35656
Y4yR Rhi	Type III secretion protein	*Rhizobium spp.*	697	2494472	spP55726
InvA Sen[c]	Type III secretion protein	*Salmonella enterica*	650	1236873	gbU43271
InvA Sty	Type III secretion protein	*Salmonella typhimurium*	665	547727	spP35657
SsaV Sty	Type III secretion protein	*Salmonella typhimurium*	681	3024658	spP74856
FlhA Sty	Flagellar biosynthesis protein	*Salmonella typhimurium*	692	729521	spP40729
MxiA Sfl	Type III secretion protein	*Shigella flexneri*	686	2506402	spP35533
Orf Sso[c]	Type III secretion protein	*Shigella sonnei*	688	829074	gbD50601
FlhA Tpa	Flagellar export protein	*Treponema pallidum*	707	1216383	gbU36839
FlhA Vpa	Flagellar biosynthesis protein	*Vibrio parahaemolyticus*	699	4322011	gbAF069392
HrcP2 Xca	Type III secretion protein	*Xanthomonas campestris*	645	462307	spP80150
YsaA Yen	Inner membrane transporter	*Yersinia enterocolitica*	690	2352530	gbAF005744
LcrD Yen	Type III secretion protein	*Yersinia enterocolitica*	704	4324341	spP21210
FlhA Yen	Flagellar biosynthesis protein	*Yersinia enterocolitica*	692	2494471	spQ56887
LcrD Ype[c]	Type III secretion protein	*Yersinia pestis*	704	400174	spP31487
LcrD Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	704	155461	gbM96850
FlhA Zmo	Flagellar biosynthesis protein	*Zymonas mobilis*	707	4378865	gbAF124349

[a]The four letter designations of the proteins are followed by three letter abbreviations of the source organism (e.g. Eco, *E. coli*). Format of presentation is the same for Tables 6-16.

[b]Accession numbers reported in this table and Tables 6-16 are derived from the following databases: GenBank, gb; Protein Information Resource, pir; and Swiss-Prot, sp.

[c]These proteins (in this table and Tables 6-16) were not included in the phylogenetic trees either because their sequences are very similar to some that were included, or because their sequences are incomplete.

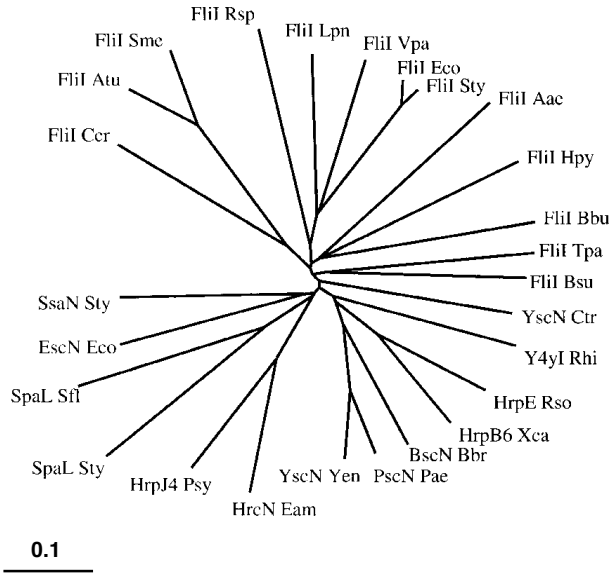


Figure 6. Phylogenetic tree for the YscN/FliI family.

(encoded in *Salmonella* pathogenicity island 1, SPI-1) and the Mxi/Spa proteins of *S. flexneri* always clustered moderately closely together. (4) The Fla proteins of *V. parahaemolyticus* cluster very loosely with the corresponding proteins of *E. coli* and *S. typhimurium* in four of six of the trees in which these proteins are all found. The two exceptions were the FliP and FliQ proteins of *V. parahaemolyticus* that did not cluster with the corresponding *E. coli* and *S. typhimurium* proteins (Figures 4C and D). (5) The *E. coli* Esc and *S. typhimurium* Ssa proteins (from the second *S. typhimurium* TIIIPS system, encoded in SPI-2) cluster loosely together in seven of the seven trees in which both homologues appear, and these two proteins cluster loosely with the Inv/Spa Sty / Mxi/Spa Sfl cluster in all of these trees. (6) All other proteins or protein clusters branch from points near the centers of these trees, and variability is probably due to experimental error. These results are consistent with the conclusion that all of the constituents of both TIIIPS and Fla systems underwent sequence divergence at fairly constant rates, and that essentially no shuffling of constituents occurred during the evolution of these systems. All of the (minor) differences between these trees may reflect experimental error. It is also important to note that phylogenetic distances for the TIIIPS systems generally do not reflect the relative phylogenetic distances of the organisms established using the sequences of 16S RNAs. However, those for the Fla proteins do represent an approximation to the phylogenetic distances observed for the organisms. It should be noted, however, that only one Gram-positive bacterium, only one Spirochete and several very distantly related Gram-negative bacteria were included in these analyses.

Table 6. YscN/FliI group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliI Atu	Flagellar-specific ATP synthase	*Agrobacterium tumefaciens*	473	2459709	gbU95165
FliI Aae	Flagellar-specific ATP synthase	*Aquifex aeolicus*	443	3913671	spO67531
FliI Bsu	Flagellar-specific ATP synthase	*Bacillus subtilis*	440	120331	spP23445
BscN Bbr	Type III secretion protein	*Bordetella bronchiseptica*	444	2935537	gbAF049488
FliI Bbu	Flagellar-specific ATP synthase	*Borrelia burgdorferi*	436	1706850	spP52607
FliI Ccr	Flagellar-specific ATP synthase	*Caulobacter crescentus*	444	1938379	gbU93180
YscN Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	442	4377010	gbAE001652
YscN Ctr	Type III secretion protein	*Chlamydia trachomatis*	442	3329120	gbAE001337
HrcN Eam	Type III secretion protein	*Erwinia amylovora*	454	1181169	gbL25828
EscN Eco	Type III secretion protein	*Escherichia coli*	446	2865290	gbAF022236
FliI Eco	Flagellar-specific ATP synthase	*Escherichia coli*	457	2506213	spP52612
FliI Hpy	Flagellar-specific ATP synthase	*Helicobacter pylori*	434	2493149	spO07025
FliI Lpn	Flagellar-specific ATP synthase	*Legionella pneumophila*	449	1938361	gbU85783
PscN Pae	Type III secretion protein	*Pseudomonas aeruginosa*	440	2459981	gbAF010151
HrpJ4 Psy	Type III secretion protein	*Pseudomonas syringae (syr)*	449	507390	gbU07346
HrpE Rso	Type III secretion protein	*Ralstonia solanacearum*	439	2120692	pirS61858
Y4yI Rhi	Type III secretion protein	*Rhizobium spp.*	451	2493150	spP55717
FliI Rsp	Flagellar-specific ATP synthase	*Rhodobacter sphaeroides*	442	1272678	gbX97201
SsaN Sty	Type III secretion protein	*Salmonella typhimurium*	433	3024659	spP74857
FliI Sty	Flagellar-specific ATP synthase	*Salmonella typhimurium*	456	120332	spP26465
SpaL Sty	Type III secretion protein	*Salmonella typhimurium*	432	730791	spP39444
SpaL Sfl	Type III secretion protein	*Shigella flexneri*	430	548969	spP35531
HrcN Sfr[c]	Type III secretion protein	*Sinorhizobium fredii*	450	1648933	gbL12251
FliI Sme	Flagellar-specific ATP synthase	*Sinorhizobium meliloti*	467	2916791	gbAJ224445
FliI Tde	Flagellar-specific ATP synthase	*Treponema denticola*	473	2105151	gbU78776
FliI Tpa	Flagellar-specific ATP synthase	*Treponema pallidum*	447	3322685	gbAE001218
FliI Vpa	Flagellar-specific ATP synthase	*Vibrio parahaemolyticus*	439	4322012	gbAF069392
HrpB6 Xca	Type III secretion protein	*Xanthomonas campestris*	442	462306	spP80153
YscN Yen	Type III secretion protein	*Yersinia enterocolitica*	439	732261	spP40290
YscN Ype[c]	Type III secretion protein	*Yersinia pestis*	439	2996241	gbAF053946
YscN Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	439	732262	spP40291

[a,b,c]footnotes for tables 6-16 are as for table 5.

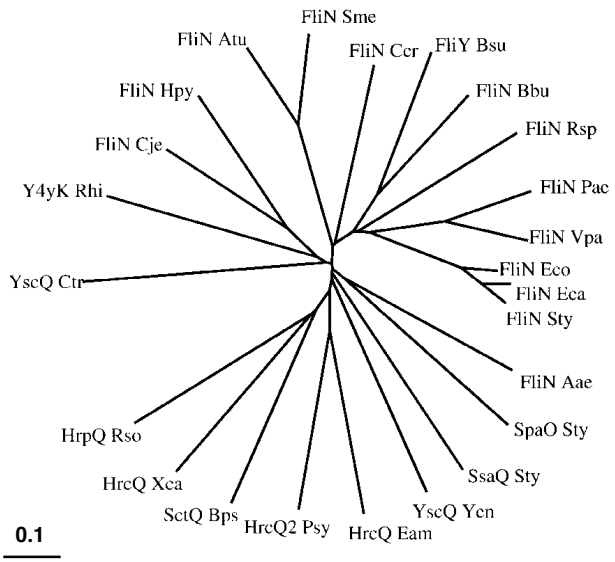


Figure 7. Phylogenetic tree for the YscQ/FliN family.

Phylogenetic Trees Constructed with All Sequenced Homologues of TIIIPS Constituents

Complete trees, including almost all homologues identified in the databases for each of the 12 families analyzed, were generated with the ClustalX program. These trees are shown in Figures 5-16. Most of the conclusions summarized above with the trees constructed using members of the complete or nearly complete TIIIPS and Fla systems (Figure 4A-F) were confirmed and extended in these trees. Therefore, only novel observations, not noted above, will be cited here.

Tables 5-16 present the members of each of the twelve families of TIIIPS protein constituents, and Figures 5-16 show the corresponding phylogenetic trees where all proteins in the tables except those noted with the superscript letter "c" were analyzed. In the tables, proteins are arranged alphabetically according to organism, and those eliminated from the reported analyses (in bold print) are those that either exhibit a high degree of similarity with another protein that was included, or are obvious fragments of incompletely or incorrectly sequenced proteins. A few of the proteins included in the phylogenetic analyses may be slightly truncated. Truncation would be expected to increase the branch length without appreciably altering the branching position. Below we systematically analyze the two principal phylogenetic clusters of these trees in greater detail.

The Flagellar (Fla) Proteins

Figures 5-11, 15 and 16 include Fla proteins. The tree in Figure 5 shows clustering of the *S. typhimurium*, *E. coli*, *Y. enterocolitica*, *P. mirabilis*, *V. parahaemolyticus* and *Z. mobilis* FlhA proteins with increasing distance in that order. Additionally, the two spirochete FlhA proteins, those of *B. burgdorferi* and *T. pallidum*, cluster loosely together as do

Table 7. YscQ/FliN Group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliN Atu	Flagellar motor switch protein	*Agrobacterium tumefaciens*	179	3913685	spQ57259
FliN Aae	Flagellar motor switch protein	*Aquifex aeolicus*	112	2983912	gbAE000745
FliY Bsu	Flagellar motor switch protein	*Bacillus subtilis*	378	585146	spP24073
FliN Bbu	Flagellar motor switch protein	*Borrelia burgdorferi*	113	2494553	spQ44903
SctQ Bps	Type III secretion protein	*Burkholderia pseudomallei*	310	4206077	gbAF074878
FliN Cje	Flagellar motor switch protein	*Campylobacter jejuni*	102	2274910	gbAJ000400
FliN Ccr	Flagellar motor switch protein	*Caulobacter crescentus*	110	462113	spQ03593
YscQ Ctr	Type III secretion protein	*Chlamydia trachomatis*	373	3329123	gbAE001337
HrcQ Eam	Type III secretion protein	*Erwinia amylovora*	338	1181172	gbL25828
FliN Eca	Flagellar motor switch protein	*Erwinia carotovora*	106	547912	spP35539
HrcQ2 Ehe[c]	Type III secretion protein	*Erwinia herbicola*	108	1483321	gbX99768
FliN Eco	Flagellar motor switch protein	*Escherichia coli*	137	120347	spP15070
FliN Hpy	Flagellar motor switch protein	*Helicobacter pylori*	123	2313688	gbAE000571
FliN Pae	Flagellar motor switch protein	*Pseudomonas aeruginosa*	157	2494554	spQ51466
HrcQ2 Psy[c]	Type III secretion protein	*Pseudomonas syringae (phs)*	128	3282780	gbAF04344
HrpU Psy	Type III secretion protein	*Pseudomonas syringae (syr)*	133	818890	gbU25812
HrpQ Rso	Type III secretion protein	*Ralstonia solanacearum*	354	2120698	pirS62086
Y4yK Rhi	Type III secretion protein	*Rhizobium spp.*	358	2494558	spP55719
FliN Rsp	Flagellar motor switch protein	*Rhodobacter sphaeroides*	152	2315250	gbY14335
SpaO Sdu[c]	Type III secretion protein	*Salmonella dublin*	303	2494555	spQ53968
SpaO Sen[c]	Type III secretion protein	*Salmonella enterica*	303	975754	gbU29359
SpaQ Sti[c]	Type III secretion protein	*Salmonella typhi*	303	2494557	spQ56022
SpaO Sty	Type III secretion protein	*Salmonella typhimurium*	303	730795	spP40699
SsaQ Sty	Type III secretion protein	*Salmonella typhimurium*	322	3024662	spP74860
FliN Sty	Flagellar motor switch protein	*Salmonella typhimurium*	137	120348	spP26419
HrcQ Sfr[c]	Type III secretion protein	*Sinorhizobium fredii*	382	1648935	gbL12251
FliN Sme	Flagellar motor switch protein	*Sinorhizobium meliloti*	198	2916786	gbAJ224445
FliN Vpa	Flagellar motor switch protein	*Vibrio parahaemolyticus*	136	4322005	gbAF069392
HrcQ Xca	Type III secretion protein	*Xanthomonas campestris*	268	4494847	gbAF056246
YscQ Yen	Type III secretion protein	*Yersinia enterocolitica*	307	4324350	gbAF102990
YscQ Ype[c]	Type III secretion protein	*Yersinia pestis*	307	1176912	spP42713
YscQ Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	307	732265	spP40296

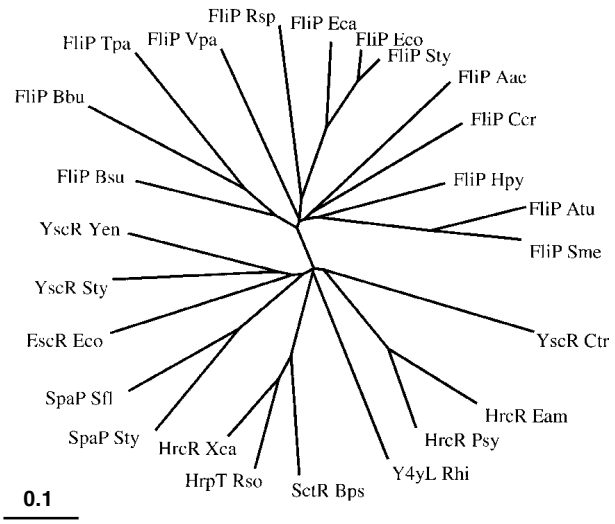


Figure 8. Phylogenetic tree for the YscR/FliP family.

the *C. jejuni* and *H. pylori* proteins. All other Fla proteins branch from points near the center of the Fla portion of the tree. These relative distances are in agreement with expectation on the basis of the phylogenetic relationships of the organisms. Because of the large sizes of these FlhA proteins (Table 5), this tree should be highly reliable. The tree for FliI homologues (Table 6; Figure 6) similarly reveals rough clustering according to the phylogenies of the organisms although the two spirochete proteins do not cluster together.

FliN homologues (Table 7; Figure 7) similarly show clustering as expected according to organismal phylogeny. As in Figures 6 and 8, *A. tumefaciens* and *S. meliloti* cluster together. These three figures are the only ones where Fla proteins from both of these organisms are represented. Contrary to expectation, the *B. burgdorferi* and *B. subtilis* proteins cluster loosely together, and no *T. pallidum* FliN protein homologue is present.

The tree for FliP homologues (Table 8; Figure 8) is essentially as expected from organismal phylogenies, except that the *R. sphaeroides* protein proves to be closer to the cluster of proteins from *S. typhimurium*, *E. coli* and *E. carotovora* than is the *V. parahaemolyticus* protein. This observation may be related to the fact that unlike the other Fla systems, that from *V. parahaemolyticus* transports Na^+ instead of H^+. As in Figure 5, the two spirochete FliP homologues cluster loosely together.

The FliQ tree (Table 9; Figure 9) is also in accordance with expectation except that the *V. parahaemolyticus* homologue clusters distantly from the *S. typhimurium-E. coli-E. carotovora* cluster as was observed for the FliP tree as noted for Figure 8. This unexpectedly great divergence between the enteric and *V. parahaemolyticus* FliP and FliQ proteins may reflect divergence for a unique function such as accommodation of Na^+ rather than H^+ as the transported ion that drives flagellar rotation. If this functional difference does prove to be responsible for the phylogenetic differences noted, it suggests that FliP and FliQ play a role in using the smf versus the pmf.

Trees for the FliR (Table 10; Figure 10), FlhB (Table 11; Figure 11), FliF (Table 15; Figure 15) and FliH (Table 16; Figure 16) families are fully consistent with expectation

Table 8. YscR/FliP Group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliP Atu	Flagellar biosynthetic protein	*Agrobacterium tumefaciens*	245	2494560	spQ44344
FliP Aae	Flagellar biosynthetic protein	*Aquifex aeolicus*	239	2984179	gbAE000763
FliP Bsu	Flagellar biosynthetic protein	*Bacillus subtilis*	221	544315	spP35528
FliP Bbu	Flagellar biosynthetic protein	*Borrelia burgdoferi*	254	2494561	spQ44763
SctR Bps	Type III secretion protein	*Burkholderia pseudomallei*	216	4206078	gbAF074878
FliP Ccr	Flagellar biosynthetic protein	*Caulobacter crescentus*	266	2494562	spQ45980
YscR Cpn	Type III secretion protein	*Chlamydia pneumoniae*	306	4377137	gbAE001663
YscR Ctr	Type III secretion protein	*Chlamydia trachomatis*	306	3329003	gbAE001327
HrcR Eam	Type III secretion protein	*Erwinia amylovora*	217	2494566	spQ46646
FliP Eca	Flagellar biosynthetic protein	*Erwinia carotovara*	258	462610	spP34200
HrcR Ehe	Type III secretion protein	*Erwinia herbicola*	217	2494567	spQ47856
EscR Eco	Type III secretion protein	*Escherichia coli*	217	2865276	gbAF022236
FliP Eco	Flagellar biosynthetic protein	*Escherichia coli*	245	416997	spP33133
FliP Hpy	Flagellar biosynthetic protein	*Helicobacter pylori*	172	4580427	gbAE000581
FliP Pae	Flagellar biosynthetic protein	*Pseudomonas aeruginosa*	141	2494563	spQ51468
FliP Ppu	Flagellar biosynthetic protein	*Pseudomonas putida*	124	4104062	gbAF031418
HrcR Psy	Type III secretion protein	*Pseudomonas syringae (phs)*	208	3282781	gbAF043444
HrpT Rso	Type III secretion protein	*Ralstonia solanacearum*	217	2494568	spQ52488
Y4yL Rhi	Type III secretion protein	*Rhizobium spp.*	222	2494571	spP55720
FliP Rsp	Flagellar biosynthetic protein	*Rhodobacter sphaeroides*	301	3435109	gbAF044580
SpaP Sen	Type III secretion protein	*Salmonella enterica*	224	973278	gbU29350
SpaP Sti	Type III secretion protein	*Salmonella typhi*	224	2494565	spQ56023
FliP Sty	Flagellar biosynthetic protein	*Salmonella typhimurium*	245	1706856	spP54700
YscR Sty	Pathogenicity islands	*Salmonella typhimurium*	215	2494570	spP74890
SpaP Sty	Type III secretion protein	*Salmonella typhimurium*	224	730796	spP40700
SpaP Sfl	Type III secretion protein	*Shigella flexneri*	216	548966	spP35529
HrcR Sfr	Type III secretion protein	*Sinorhizobium fredii*	249	1648936	gbL12251
FliP Sme	Flagellar biosynthetic protein	*Sinorhizobium meliloti*	245	2506424	spP37827
FliP Tde	Flagellar biosynthetic protein	*Treponema denticola*	271	4426951	gbAF122909
FliP Tpa	Flagellar biosynthetic protein	*Treponema pallidum*	271	2494564	spP74930
FliP Vpa	Flagellar biosynthetic protein	*Vibrio parahaemolyticus*	289	4322007	gbAF069392
HrcR Xca	Pathogenicity island	*Xanthomonas campestris*	216	1346805	spP37828
YscR Yen	Type III secretion protein	*Yersinia enterocolitica*	217	4324351	gbAF102990
YscR Ype	Type III secretion protein	*Yersinia pestis*	217	732266	spP40297
YscR Yps	Type III secretion protein	*Yersinia pseudotuberculosis*	217	475123	gbL25667

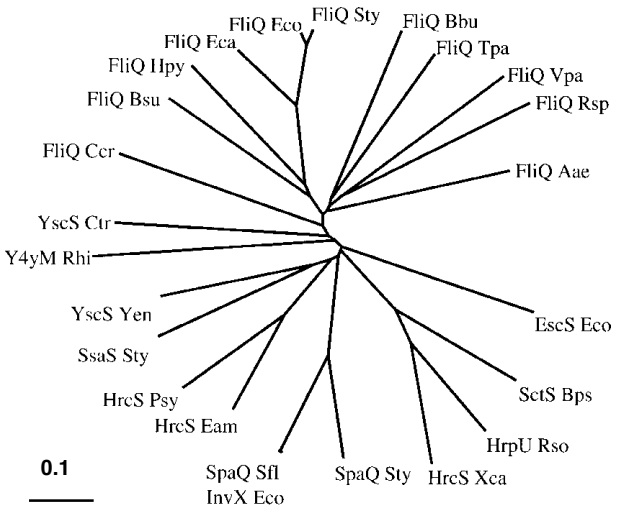


Figure 9. Phylogenetic tree for the YscS/FliQ family.

based on organismal phylogenies. Moreover, clustering of the two distantly related spirochete proteins is as expected for six of the eight trees in which both proteins are represented. The exceptions were FliI (Figure 6) and FliR (Figure 10). It can therefore be concluded that with only a very few exceptions, all 9 Fla protein families exhibit clustering patterns in agreement with organismal phylogeny, a result consistent with the suggestion that all or most of the Fla homologues portrayed in these trees are orthologues. In these cases, the exceptions noted may be due to experimental error.

The Type III Protein Secretion (TIIIPS) Proteins

As noted above and illustrated in the trees depicted in Figure 4, several pairs of close homologues derived from the same TIIIPS systems paired together in all trees represented. The more complete trees shown in Figures 5-16 confirmed and extended this conclusion. For example, all *R. solanacearum* proteins are most closely related to the corresponding homologues from *X. campestris*, and all but one of the *Y. enterocolitica* proteins are most similar to the homologues in *P. aeruginosa* when the sequences of these proteins are available. Further, the *E. amylovora* proteins paired with either the *P. syringae* (phs) proteins (Figures 8-11) or the *P. syringae* (syr) proteins (Figures 6, 7 and 12). The systems from these two strains of *P. syringae* are probably very similar. However, one of the trees (Figure 5) proved exceptional. The two exceptions shown in Figure 5 were HrpI of *E. amylovora* that clustered with PcrD of *P. aeruginosa*, and LcrD of *Y. enterocolitica* that paired with HrpI of *P. syringae* (syr). The *P. syringae* (phs) LcrD homologue is not available. Note: except for HrpI of *P. syringae* (Figure 5) all other *P. aeruginosa* proteins paired with the *Y. enterocolitica* proteins in the trees shown in Figures 12-16. Thus, this one example (Figure 5) represents the single case in which horizontal transfer of

Table 9. YscS/FliQ Group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliQ Aae	Flagellar biosynthetic protein	*Aquifex aeolicus*	89	2984201	gbAE000765
FliQ Bsu	Flagellar biosynthetic protein	*Bacillus subtilis*	89	544316	spP35535
FliQ Bbu	Flagellar biosynthetic protein	*Borrelia burgdorferi*	87	3023774	spQ44906
SctS Bps	Type III secretion protein	*Burkholderia pseudomallei*	87	4206079	gbAF074878
FliQ Ccr	Flagellar biosynthetic protein	*Caulobacter crescentus*	87	3023776	spQ45974
YscS Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	95	4377136	gbAE001663
YscS Ctr	Type III secretion protein	*Chlamydia trachomatis*	94	3329004	gbAE001327
HrcS Eam	Type III secretion protein	*Erwinia amylovora*	86	1181174	gbL25828
FliQ Eca	Flagellar biosynthetic protein	*Erwinia carotovora*	89	462611	spP34201
HrcS Ehe[c]	Type III secretion protein	*Erwinia herbicola*	76	1483323	gbX99768
EscS Eco	Type III secretion protein	*Escherichia coli*	89	2865277	gbAF022236
FliQ Eco	Flagellar biosynthetic protein	*Escherichia coli*	89	416998	spP33134
InvX Eco	Invasion membrane protein	*Escherichia coli*	86	538746	pirA40611
FliQ Hpy	Flagellar biosynthetic protein	*Helicobacter pylori*	88	2314593	gbAE000642
HrcS Psy	Type III secretion protein	*Pseudomonas syringae (phs)*	88	3282782	gbAF043444
HrpU Rso	Type III secretion protein	*Ralstonia solanacearum*	86	2120700	pirS61850
Y4yM Rhi	Type III secretion protein	*Rhizobium spp.*	91	2494572	spP55721
FliQ Rsp	Flagellar biosynthetic protein	*Rhodobacter sphaeroides*	88	3435110	gbAF044580
SpaQ Sdu[c]	Type III secretion protein	*Salmonella dublin*	86	973259	gbU29345
SpaQ Sen[c]	Type III secretion protein	*Salmonella enterica*	86	973295	gbU29354
SsaS Sty	Type III secretion protein	*Salmonella typhimurium*	88	3024663	spP74891
SpaQ Sty	Type III secretion protein	*Salmonella typhimurium*	82	730797	spP40704
FliQ Sty	Flagellar biosynthetic protein	*Salmonella typhimurium*	89	1706857	spP54701
SpaQ Sfl	Type III secretion protein	*Shigella flexneri*	86	730798	spP40705
Orf Sso[c]	Type III secretion protein	*Shigella sonnei*	80	829081	gbD50601
HrcS Sfr[c]	Type III secretion protein	*Sinorhizobium fredii*	92	1648937	gbL12251
FliQ Tpa	Flagellar biosynthetic protein	*Treponema pallidum*	94	3023764	spP74931
FliQ Vpa	Flagellar biosynthetic protein	*Vibrio parahaemolyticus*	89	4322008	gbAF069392
HrcS Xca	Type III secretion protein	*Xanthomonas campestris*	86	4494849	gbAF056246
YscS Yen	Type III secretion protein	*Yersinia enterocolitica*	88	4324352	gbAF102990
YscS Ype[c]	Type III secretion protein	*Yersinia pestis*	79	1176913	spP42715
YscS Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	88	732267	spP40298

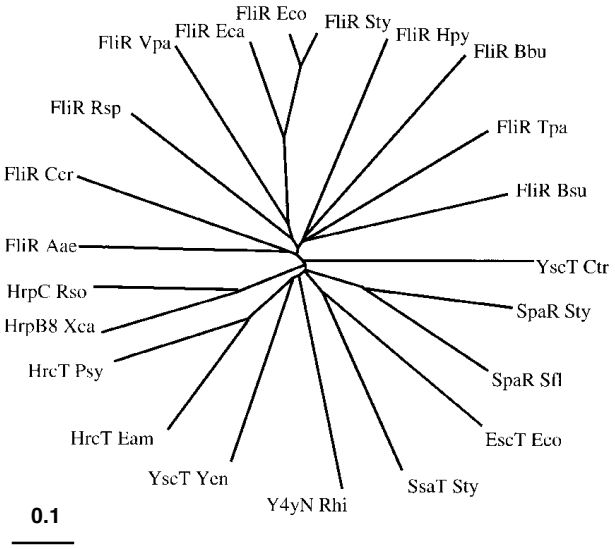


Figure 10. Phylogenetic tree for the YscT/FliR family.

the gene *pcrD* of *P. aeruginosa* might have occurred from a system similar to the *E. amylovora* system, replacing the homologue which should be similar to the protein from *Y. enterocolitica*.

The Secretin Phylogenetic Tree

The tree of secretins shown in Figure 12 shows that all TIIIPS secretins cluster together on a single branch. The same is observed for the phage-encoded secretins. Further, all of the general Type II protein secretion (TIIPS) secretins cluster loosely together. The same is true of the pilin secretins. However, the ComE competence protein of *H. influenzae* clusters loosely with the pilin secretins. Finally, the NolW protein of *Rhizobium*, which plays an unidentified biochemical role in nodulation of the host plant by the bacterium, does not cluster with any of the other secretins, branching from a point near the center of the tree.

Discussion

Type III protein secretion (TIIIPS) systems are found exclusively in Gram-negative bacterial pathogens and may facilitate transfer of virulence proteins directly from the bacterial cytoplasm to the eukaryotic host cell cytoplasm in a single energy-coupled step (Lee, 1997; Hueck, 1998; Galán and Collmer, 1999). Although it seems clear that gene clusters encoding these systems have been transferred horizontally between various pathogens (Groisman and Ochman, 1993; Lee, 1996; Groisman and Ochman, 1997), their evolutionary origins and the mechanisms of their transfer are poorly understood (Karaolis *et al.,* 1999; Mirold *et al.,* 1999).

The gene clusters which encode the TIIIPS systems usually comprise genes for secreted virulence proteins that produce pathogenic responses in the host animal or plant cell (reviewed in Hueck, 1998). In contrast to the sequence conservation of the majority of constituents of the TIIIPS apparatus, the secreted virulence proteins are highly variable both within a given pathogen and between pathogens. Many of these virulence proteins show structural and/or biochemical similarity to certain signal transduction proteins of the host cell and interfere with host cell signal transduction pathways (Guan and Dixon, 1990; Rosqvist *et al.,* 1991; Galyov *et al.,* 1993; Collmer and Bauer, 1994; Bonas, 1994; Chen *et al.,* 1996; Hardt *et al.,* 1998; Mudgett and Staskawicz, 1998; Norris *et al.,* 1998; Fu and Galán, 1998; Spiik *et al.,* 1999). Therefore, the genes encoding the secreted proteins may have been captured from the host by the pathogen during the evolutionary past. Since specificity of secretion appears to

Table 10. YscT/FliR group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliR Aae	Flagellar biosynthetic protein	*Aquifex aeolicus*	258	2984202	gbAE000765
FliR Bsu	Flagellar biosynthetic protein	*Bacillus subtilis*	259	544317	spP35537
FliR Bbu	Flagellar biosynthetic protein	*Borrelia burgdorferi*	269	3913681	spQ44907
FliR Ccr	Flagellar biosynthetic protein	*Caulobacter crescentus*	251	3023777	spQ45975
YscT Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	289	4377135	gbAE001663
YscT Ctr	Type III secretion protein	*Chlamydia trachomatis*	289	3329005	gbAE001327
HrcT Eam	Type III secretion protein	*Erwinia amylovora*	265	1181175	gbL25828
FliR Eca	Flagellar biosynthetic protein	*Erwinia carotovora*	261	462612	spP34202
HrcT Ehe[c]	Type III secretion protein	*Erwinia herbicola*	265	1483324	gbX99768
EscT Eco	Type III secretion protein	*Escherichia coli*	258	2865278	gbAF022236
FliR Eco	Flagellar biosynthetic protein	*Escherichia coli*	261	2506425	spP33135
FliR Hpy	Flagellar biosynthetic protein	*Helicobacter pylori*	255	2313256	gbAE000537
FliR Ppu[c]	Flagellar biosynthetic protein	*Pseudomonas putida*	130	4104063	gbAF031418
HrcT Psy	Type III secretion protein	*Pseudomonas syringae (phs)*	264	3282783	gbAF043444
HrcC Rso	Type III secretion protein	*Ralstonia solanacearum*	282	2120690	pirS61860
Y4yN Rhi	Type III secretion protein	*Rhizobium spp.*	272	2494573	spP55722
FliR Rsp	Flagellar biosynthetic protein	*Rhodobacter sphaeroides*	269	3435111	gbAF044580
FliR Sty	Flagellar biosynthetic protein	*Salmonella typhimurium*	264	1706858	spP54702
SsaT Sty	Type III secretion protein	*Salmonella typhimurium*	259	3024667	spP96068
SpaR Sty	Type III secretion protein	*Salmonella typhimurium*	263	730799	spP40701
SpaR Sfl	Type III secretion protein	*Shigella flexneri*	256	730800	spP40706
HrcT Sfr[c]	Type III secretion protein	*Sinorhizobium fredii*	272	1648938	gbL12251
FliR Tpa	Flagellar biosynthetic protein	*Treponema pallidum*	265	3023765	spP74932
FliR Vpa	Flagellar biosynthetic protein	*Vibrio parahaemolyticus*	260	4322009	gbAF069392
HrpB8 Xca	Type III secretion protein	*Xanthomonas campestris*	276	984885	gbU33548
YscT Yen	Type III secretion protein	*Yersinia enterocolitica*	261	4324353	gbAF102990
YscT Ype[c]	Type III secretion protein	*Yersinia pestis*	261	3822065	gbAF074612
YscT Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	261	732268	spP40299

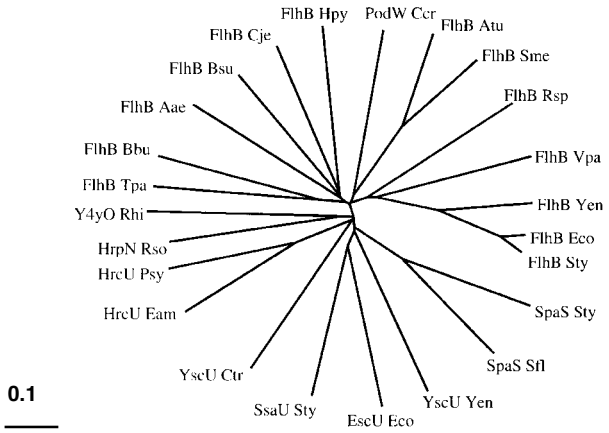


Figure 11. Phylogenetic tree for the YscU/FlhB family.

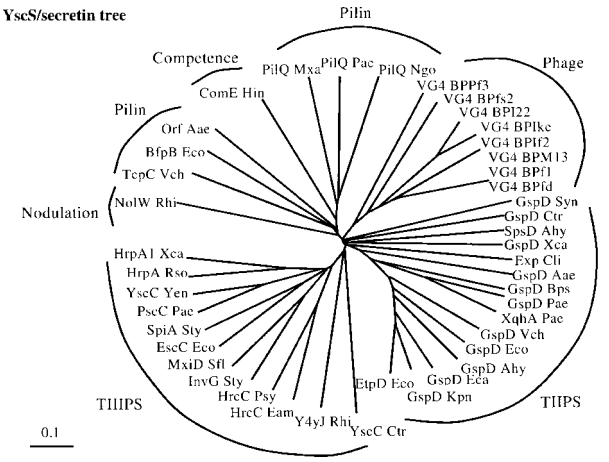


Figure 12. Phylogenetic tree for the YscC/secretin family.

be dictated by the secreted proteins (and possibly cognate chaperonic factors; Cheng and Schneewind, 1999) rather than by biochemical properties of the secretory system (Hermant *et al.,* 1995; Rosqvist *et al.,* 1995; Frithz-Lindsten *et al.,* 1997; Anderson *et al.,* 1999; Rossier *et al.,* 1999), TIIIPS apparatuses appear to function as multi-purpose bacterial syringes, capable of injecting a variety of different poisons into host cells. While the biochemical functions of many virulence proteins secreted by TIIIPS systems have been analyzed in great detail, very little is known about the biochemistry of the actual secretion process. On the contrary, abundant sequence information for TIIIPS system genes is available. Here we have used this information to analyze the evolutionary relationships of the constituents of TIIIPS systems. We have used two programs, and analyses were conducted either (a) with just those

Table 11. YscU/FlhB group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FlhB Atu	Flagellar biosynthetic protein	*Agrobacterium tumefaciens*	360	2459702	gbU95165
FlhB Aae	Flagellar biosynthetic protein	*Aquifex aeolicus*	350	2984250	gbAE000768
FlhB Bsu	Flagellar biosynthetic protein	*Bacillus subtilis*	360	544313	spP35538
FlhB Bbu	Flagellar biosynthetic protein	*Borrelia burgdorferi*	372	3023773	spQ44760
FlhB Cje	Flagellar biosynthetic protein	*Campylobacter jejuni*	362	4105266	gbAF044271
PodW Ccr	Flagellar biosynthetic protein	*Caulobacter crescentus*	361	1147737	gbU42203
Cds1 Cca[c]	Type III secretion protein	*Chlamydia caviae*	360	2444073	gbU88070
YscU Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	360	4376600	gbAE001617
YscU Ctr	Type III secretion protein	*Chlamydia trachomatis*	360	3328487	gbAE001283
HrcU Eam	Type III secretion protein	*Erwinia amylovora*	360	1181176	gbL25828
HrcU Ehe[c]	Flagellar biosynthetic protein	*Erwinia herbicola*	52	1483325	gbX99768
EscU Eco	Type III secretion protein	*Escherichia coli*	345	2865279	gbAF022236
FlhB Eco	Flagellar biosynthetic protein	*Escherichia coli*	382	2494574	spP76299
FlhB Hpy	Flagellar biosynthetic protein	*Helicobacter pylori*	358	3023763	spP56416
FlhB Pmi	Flagellar biosynthetic protein	*Proteus mirabilis*	74	2126141	pirS61501
HrcU Psy[c]	Type III secretion protein	*Pseudomonas syringae (gly)*	325	3603323	gbAF069652
HrcU Psy	Type III secretion protein	*Pseudomonas syringae (phs)*	359	3282784	gbAF043444
HrpY Psy[c]	Type III secretion protein	*Pseudomonas syringae (syr)*	359	818894	gbU25812
HrpN Rso	Type III secretion protein	*Ralstonia solanacearum*	357	547676	spP35652
Y4yO Rhi	Type III secretion protein	*Rhizobium spp.*	345	2494576	spP55723
FlhB Rsp	Flagellar biosynthetic protein	*Rhodobacter sphaeroides*	376	3435112	gbAF044580
SsaU Sty	Type III secretion protein	*Salmonella typhimurium*	352	3024668	spP96069
SpaS Sty	Type III secretion protein	*Salmonella typhimurium*	356	730801	spP40702
FlhB Sty	Flagellar biosynthetic protein	*Salmonella typhimurium*	383	729522	spP40727
SpaS Sfl	Type III secretion protein	*Shigella flexneri*	342	730802	spP40707
HrcU Sfr[c]	Type III secretion protein	*Sinorhizobium fredii*	351	1648939	gbL12251
FlhB Sme	Flagellar biosynthetic protein	*Sinorhizobium meliloti*	360	2916784	gbAJ224445
FlhB Tpa	Flagellar biosynthetic protein	*Treponema pallidum*	376	3323017	gbAE001244
FlhB Vpa	Flagellar biosynthetic protein	*Vibrio parahaemolyticus*	376	4322010	gbAF069392
FlhB Yen	Flagellar biosynthetic protein	*Yersinia enterocolitica*	383	2494575	spQ56886
YscU Yen	Type III secretion protein	*Yersinia enterocolitica*	354	4324354	gbAF102990
YscU Ype[c]	Type III secretion protein	*Yersinia pestis*	354	3822064	gbAF074612
YscU Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	354	732269	spP40300

homologues for which completely (or nearly completely) sequenced TIIIPS systems are available (Figures 4A-F) or (b) with all (or most) available homologues (Figures 5-16). The two programs used gave comparable results, and regardless of the approach used, the results led to similar conclusions. Several of these conclusions are cited below and interpreted in terms of an evolutionary scheme for the appearance and maintenance of distinct TIIIPS systems.

The homologous protein constituents of different virulence-related TIIIPS systems cluster similarly (*i.e.*, in accordance with system phylogeny) in all trees that include sequenced homologues of a particular TIIIPS protein type. This notion is true for the inner membrane components which are homologous to flagellar proteins and for the outer membrane secretins. The pairwise clustering of protein systems corresponds to the pairwise similarities in genetic organization of the encoding genes (see Figure 1). The phylogenetic relationships of these TIIIPS proteins clearly do not follow those of the source organisms (see, for example, the 16S rRNA-derived organismal differences between *R. solanacearum* and *X. campestris*, two pathogens which carry very closely related TIIIPS systems).

Table 12. YscC/secretin group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
GspD Ahy	General secretion protein	*Aeromonas hydrophila*	678	1170050	spP31780
SpsD Ahy	S-protein secretion protein	*Aeromonas hydrophila*	737	2126227	pirl39547
GspD Asa[c]	General secretion protein	*Aeromonas salmoncida*	678	1170051	spP45778
GspD Aae	General secretion protein	*Aquifex aeolicus*	625	2983708	gbAE000732
Orf Aae	Pilin transport protein	*Aquifex aeolicus*	705	2983222	gbAE000697
VG4 BPf1	Gene IV protein secretion	*Bacteriophage f1*	426	138046	spP03666
VG4 BPfd	Gene IV protein secretion	*Bacteriophage fd*	426	138047	spP03664
VG4 BPfs2	Gene IV protein secretion	*Bacteriophage fs-2*	500	3702216	gbAB002632
VG4 BPI22	Gene IV protein secretion	*Bacteriophage I2-2*	429	138048	spP15420
VG4 BPIf1	Gene IV protein secretion	*Bacteriophage If1*	429	3676288	gbU02303
VG4 BPIke	Gene IV protein secretion	*Bacteriophage Ike*	437	138049	spP03667
VG4 BPM13	Gene IV protein secretion	*Bacteriophage M13*	426	138050	spP03665
VG4 BPPf3	Gene IV protein secretion	*Bacteriophage Pf3*	430	138034	spP03668
GspD Bps	General secretion protein	Burkholderia pseudomallei	750	4139236	gbAF110185
YscC Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	919	4377005	gbAE001652
GspD Cpn[c]	General secretion protein	*Chlamydia pneumoniae*	754	4377127	gbAE001662
YscC Ctr	Type III secretion protein	*Chlamydia trachomatis*	921	3329125	gbAE001337
GspD Ctr	General secretion protein	*Chlamydia trachomatis*	760	3329013	gbAE001327
Exp Cli	Exporter protein	*Chlorobium limicola*	461	1688247	gbU77780
Orf Cbu[c]	Hypothetical protein	*Coxiella burnetii*	171	3248946	gbAF06949
HrcC Eam	Type III secretion protein	*Erwinia amylovora*	676	1336093	gbU56662
GspD Eca	General secretion protein	*Erwinia carotovora*	659	2506491	spP31701
HrcC Ech[c]	Type III secretion protein	*Erwinia chrysanthemi*	691	1772618	gbL39897
GspD Ech[c]	General secretion protein	*Erwinia chrysanthemi*	712	399792	spP31700
EscC Eco	Type III secretion protein	*Escherichia coli*	512	2897962	gbAF022236
GspD Eco	General secretion protein	*Escherichia coli*	654	1170052	spP45758
EtpD Eco	Type II secretion protein	*Escherichia coli*	585	2598401	gbY09824
HofQ Eco[c]	Outer membrane transport	*Escherichia coli*	412	1170332	spP34749
BfpB Eco	Bundle-forming pilus protein	*Escherichia coli*	552	1314252	gbU27184
ComE Hin	Outer membrane transport	*Haemophilus influenzae*	445	1169008	spP31772
GspD Kpn	General secretion protein	*Klebsiella pneumoniae*	660	131592	spP15644
PilQ Mxa	Pilin transport protein	*Myxococcus xanthus*	901	3978519	gbAF100157
PilQ Ngo	Pilin transport protein	*Neisseria gonorrhoeae*	720	2120880	pirS70838
PilQ Nme[c]	Pilin transport protein	*Neisseria meningitidis*	766	4027986	gbAF066056
PscC Pae	Type III secretion protein	*Pseudomonas aeruginosa*	600	1781385	gbU56077
GspD Pae	General secretion protein	*Pseudomonas aeruginosa*	658	544439	spP35818
XqhA Pae	Type II secretion protein	*Pseudomonas aeruginosa*	713	2853000	gbAF044261
PilQ Pae	Pilin transport protein	*Pseudomonas aeruginosa*	714	464392	spP34750
XcpQ Pal[c]	Type II secretion protein	*Pseudomonas alcaligenes*	649	3978475	gbAF092918
XcpQ Ppu[c]	Type II secretion protein	*Pseudomonas putida*	591	2120685	pirS64727
HrcC Psy	Type III secretion protein	*Pseudomonas syringae (syr)*	701	2934881	gbL01064
HrpA Rso	Type III secretion protein	*Ralstonia solanacearum*	568	2833448	spQ52498
Y4xJ Rhi	Type III secretion protein	*Rhizobium spp.*	423	2495099	spP55702
NolW Rhi	Nodulation protein	*Rhizobium spp.*	234	2495098	spP55712
SpiA Sty	Type III secretion protein	*Salmonella typhimurium*	497	1498307	gbU51927
InvG Sty	Type III secretion protein	*Salmonella typhimurium*	562	1170574	spP35672
MxiD Sfl	Type III secretion protein	*Shigella flexneri*	566	547951	spQ04641
MxiD Sso[c]	Type III secretion protein	*Shigella sonnei*	566	2495097	spQ55293
NolW Sfr[c]	Nodulation protein	*Sinorhizobium fredii*	234	462733	spP33212
GspD Syn	General secretion protein	*Synechocystis spp.*	785	1653364	gbD90913
GspD Vch	General secretion protein	*Vibrio cholerae*	674	1170053	spP45779
TcpC Vch	Pilin transport protein	*Vibrio cholerae*	489	267086	spP29481
GspD Xca	General secretion protein	*Xanthomonas campestris*	759	129748	spP29041
HrpA1 Xca	Type III secretion protein	*Xanthomonas campestris*	607	462304	spP80151
HrpA Xor[c]	Type III secretion protein	*Xanthomonas oryzae*	37	4003501	gbAF026197
YscC Yen	Type III secretion protein	*Yersinia enterocolitica*	607	4324359	spQ01244
YscC Ype[c]	Type III secretion protein	*Yersinia pestis*	607	3822061	gbAF074612

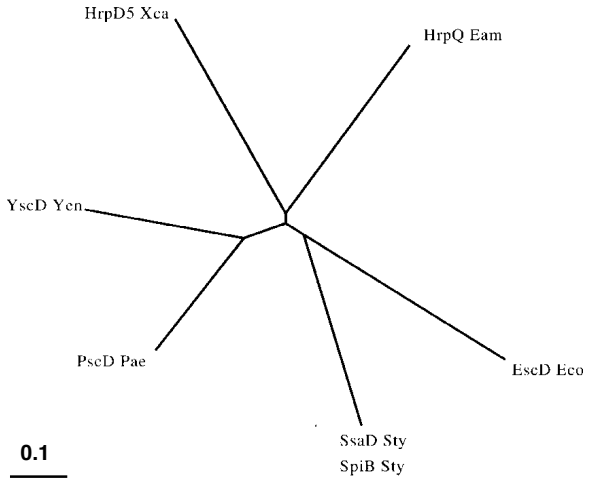


Figure 13. Phylogenetic tree for the YscD family.

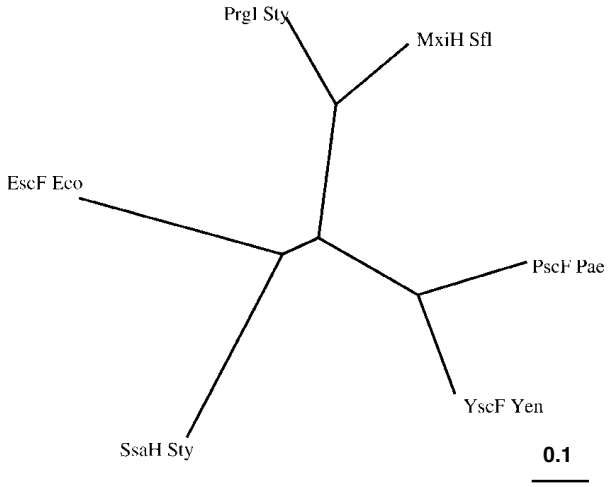


Figure 14. Phylogenetic tree for the YscF family.

Table 13. YscD Family

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
HrpQ Eam	Type III secretion protein	*Erwinia amylovora*	310	1181168	gbL25828
EscD Eco	Type III secretion protein	*Escherichia coli*	406	2897964	gbAF022236
PscD Pae	Type III secretion protein	*Pseudomonas aeruginosa*	432	1781386	gbU56077
SpiB Sty	Type III secretion protein	*Salmonella typhimurium*	323	1498308	gbU51927
SsaD Sty	Type III secretion protein	*Salmonella typhimurium*	231	3776116	gbAJ224892
HrpD5 Xca	Type III secretion protein	*Xanthomonas campestris*	312	4494852	gbAF056246
YscD Yen	Type III secretion protein	*Yersinia enterocolitica*	418	4324360	spQ01245
YscD Ype[c]	Type III secretion protein	*Yersinia pestis*	419	155447	gbM83225

Table 14. YscF Family

Designation[a]	Description	Organisim	No. of residues	GenBank GI No.	Accession No.[b]
EscF Eco	Type III secretion protein	*Escherichia coli*	73	1929439	gbL76581
PscF Pae	Type III secretion protein	*Pseudomonas aeruginosa*	85	1781388	gbU56077
PrgI Sty	Type III secretion protein	*Salmonella typhimurium*	80	1172613	spP41784
SsaH Sty	Type III secretion protein	*Salmonella typhimurium*	71	2460267	gbAF020808
MxiH Sfl	Type III secretion protein	*Shigella flexneri*	83	547953	spQ06079
YscF Yen	Type III secretion protein	*Yersinia enterocolitica*	87	267567	spQ01247
YscF Ype[c]	Type III secretion protein	*Yersinia pestis*	87	2996226	gbAF053946

Thus, lateral transfer of the genetic clusters encoding TIIIPS systems between Gram-negative pathogens has occurred repeatedly. Furthermore, the phylogenetic analyses confirm that such systems have virtually always been transferred intact without the formation of hybrid systems, even when more than one such system is present in a single bacterium (as for the *S. typhimurium* TIIIPS systems Inv/Spa and Ssa). Horizontal transfer of intact TIIIPS systems may have been facilitated by the presence of repeat sequences flanking the genetic apparatuses encoding these systems (Hueck, 1998) as well as by use of bacteriophage as DNA carriers (see Karaolis *et al.,* 1999 for illustration of such potential occurrences). Repeat sequences would allow genetic recombination, promoting integration and removal of the encompassed circular DNA regions, just as plasmids and lysogenic phage integrate into and excise from bacterial chromosomes (Buchrieser *et al.,* 1998; Franco *et al.,* 1999). Such a mechanism would facilitate transfer of an entire TIIIPS system without introducing the possibility of hybrid system formation as noted above. Dependency of the intact system on numerous system-specific protein-protein interactions would further insure the lack of hybrid system formation. The more divergent in sequence two systems are, the greater the restriction to hybrid formation would be expected to be. If two systems were very similar in sequence, the formation of hybrid systems would

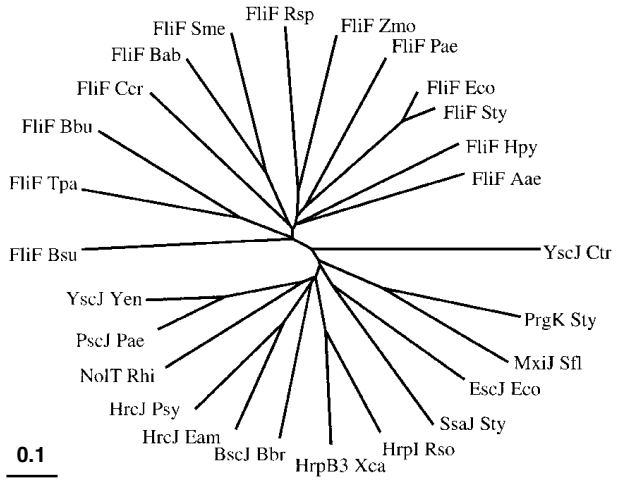


Figure 15. Phylogenetic tree for the YscJ/FliF family.

theoretically be possible without appreciable loss of activity, but phylogenetic analyses would not allow detection of such an event. Distinct regulatory constraints imposed upon two different systems within a single cell would tend to further prohibit hybrid interactions.

The nearly complete absence of lateral exchange of single TIIIPS constituents between different systems suggests that individual TIIIPS systems have evolved as biochemically integrated units. Thus, we conclude that the secretion process occurs in a single cooperative step which relies on extensive structural and functional protein-protein interactions between inner and outer membrance components within the TIIIPS apparatus. Furthermore, the actual biochemical interactions between constituents of the secretory apparatus may differ between different TIIIPS systems. This notion is further supported by the fact that all TIIIPS systems comprise secretory genes which are not broadly conserved in other systems (see Figure 1).

Flagellar proteins that are homologous to TIIIPS system constituents cluster separately from the TIIIPS proteins. It can thus be concluded that the TIIIPS and Fla systems diverged from each other from a single primordial system very early, before divergence of each of these two types of systems occurred. In contrast to the TIIIPS constituents, the flagellar constituents do reflect the phylogenies of the source organisms, demonstrating that individual flagellar systems diverged predominantly by speciation while TIIIPS systems appear to have their evolutionary origin in duplication of the genetic apparatuses encoding these systems. The notion of subsequently independent evolution of the flagellar and TIIIPS systems is further supported by the fact that not all constituents of TIIIPS systems are found in the bacterial flagellum and vice versa (see Figure 1 and Table 2).

We suggest that the flagellar basal body was the evolutionary progenitor of the inner membrane components of the virulence-related TIIIPS system. The early appearance of flagelli is consistent with the facts that flagelli are found in both Gram-positive and Gram-negative

Table 15. YscJ/FliF Group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliF Aae	Flagellar M-ring protein	*Aquifex aeolicus*	289	2983628	gbAE000727
FliF Bsu	Flagellar M-ring protein	*Bacillus subtilis*	536	120324	spP23447
BscJ Bbr	Type III secretion protein	*Bordetella bronchiseptica*	274	2935534	gbAF049488
FliF Bbu	Flagellar M-ring protein	*Borrelia burgdorferi*	569	3915695	spQ44912
FliF Bab	Flagellar M-ring protein	*Brucella abortus*	540	2832233	gbAF019251
FliF Ccr	Flagellar M-ring protein	*Caulobacter crescentus*	536	462109	spQ04954
YscJ Cpn[c]	Type III secretion protein	Chlamydia pneumoniae	335	4377140	gbAE001663
YscJ Ctr	Type III secretion protein	*Chlamydia trachomatis*	326	3329000	gbAE001327
HrcJ Eam	Type III secretion protein	Erwinia amylovora	260	1336088	gbU56662
EscJ Eco	Type III secretion protein	*Escherichia coli*	190	2865286	gbAF022236
FliF Eco	Flagellar M-ring protein	*Escherichia coli*	552	2506421	spP25798
FliF Hpy	Flagellar M-ring protein	*Helicobacter pylori*	567	4154846	gbAE001468
PscJ Pae	Type III secretion protein	*Pseudomonas aeruginosa*	248	1781392	gbU56077
FliF Pae	Flagellar M-ring protein	*Pseudomonas aeruginosa*	599	2494548	spQ51463
HrcJ Psy	Type III secretion protein	*Pseudomonas syringae (gly)*	268	3603315	gbAF069651
HrpC Psy[c]	Type III secretion protein	*Pseudomonas syringae (syr)*	268	818896	gbU25813
HrpI Rso	Type III secretion protein	*Ralstonia solanacearum*	269	2120695	pirS61855
NolT Rhi	Nodulation protein	*Rhizobium spp.*	289	2497732	spP55714
FliF Rsp	Flagellar M-ring protein	*Rhodobacter sphaeroides*	570	2494549	spQ53151
SsaJ Sty	Type III secretion protein	*Salmonella typhimurium*	249	3024642	spP74852
PrgK Sty	Type III secretion protein	*Salmonella typhimurium*	252	1172615	spP41786
FliF Sty	Flagellar M-ring protein	*Salmonella typhimurium*	560	120326	spP15928
MxiJ Sfl	Type III secretion protein	*Shigella flexneri*	241	547955	spQ06081
MxiJ Sso[c]	Type III secretion protein	*Shigella sonnei*	241	2497731	spQ55288
NolT Sfr[c]	Nodulation protein	*Sinorhizobium fredii*	289	462730	spP33209
FliF Sme	Flagellar M-ring protein	*Sinorhizobium meliloti*	557	2916780	gbAJ224445
FliF Tde[c]	Flagellar M-ring protein	*Treponema denticola*	567	2105148	gbU78776
FliF Tpa	Flagellar M-ring protein	*Treponema pallidum*	567	3322682	gbAE001218
HrpB3 Xca	Type III secretion protein	*Xanthomonas campestris*	253	462305	spP80152
YscJ Yen	Type III secretion protein	*Yersinia enterocolitica*	244	267571	spQ01251
YscJ Ype[c]	Type III secretion protein	*Yersinia pestis*	236	3822055	gbAF074612
YscJ Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	244	401692	spQ00926
FliF Zmo	Flagellar M-ring protein	*Zymomonas mobilis*	555	4378873	gbAF124349

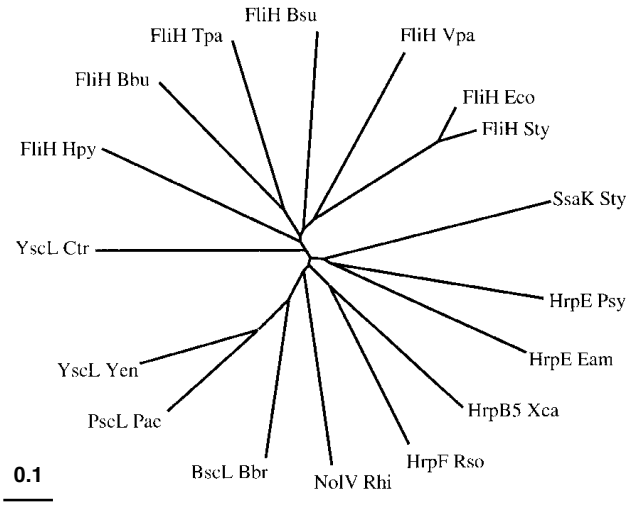


Figure 16. Phylogenetic tree for the YscL/FliH family.

bacteria (although not in archaea or eukaryotes) and that the homologous constituents of these organelles exhibit phylogenetic relationships that approximate those of the source organisms. The later appearance of virulence-related TIIIPS systems from the flagellar apparatus of a Gram-negative bacterium is consistent with (1) the restriction of virulence-related TIIIPS systems to Gram-negative bacteria, (2) the assumption that TIIIPS systems were selected by parasitic or symbiotic interactions with eukaryotes, and (3) the fact that higher eukaryotes appeared only within the last billion years, long after the appearance of - probably flagellated - prokaryotes, about 4 billion years ago. This causal evolutionary relationship has recently been suggested independently in a brief commentary by MacNab (MacNab, 1999). However, the structural and potential functional homologies between the flagella export apparatus and TIIIPS systems are restricted to inner membrane components. The constitutive presence of the outer membrane secretins in TIIIPS systems (see below) clearly distinguishes TIIIPS systems from the flagellar export apparatus. Thus, since TIIIPS systems apparently function as integrated units, we do not agree with the notion that "the flagella pathway is a type III pathway" (MacNab, 1999). This argument is further supported by the fact that virulence-related TIIIPS systems are apparently restricted to Gram-negative bacteria, while the homologous flagellar proteins are found in both Gram-negative and Gram-positive bacteria.

The absence of TIIIPS systems from Gram-positive bacteria may appear surprising at first sight, since lateral transfer and fixation of antibiotic resistance genes has occurred repeatedly between Gram-negative and Gram-positive bacteria, and the acquisition of a TIIIPS system could have provided extreme evolutionary advantages for symbiotic and parasitic bacteria, *e.g.* by facilitating escape from immune recognition due to invasion of eukaryotic cells. We rationalize the restriction of TIIIPS systems to Gram-negative bacteria by assuming that these complex systems evolved for the purpose of protein secretion across three membranes of which the Gram-negative bacterial outer membrane is the crucial one. In the absence of an outer membrane, the outer membrane constituents of TIIIPS systems become superfluous, and in fact cannot be accommodated. Thus, any protein-protein interactions between inner membrane constituents and outer membrane constituents as well as between outer membrane constituents and host eukaryotic cell membrane constituents must be abolished if a type TIIIPS system is to function in a Gram-positive bacterium, and totally new types of interactions must be established. Such modifications to the system might require extensive periods of evolutionary time and would clearly provide a basis for restricting lateral transfer of TIIIPS systems to Gram-positive bacteria, particularly if protein-protein interactions in the TIIIPS systems are cooperative. In this regard, it is relevant to note that all TIIIPS systems characterized to date include an outer membrane secretin. If the secretin is essential to overall function, the constitutive presence of secretins in TIIIPS systems provides a clue as to why type III protein secretion has not been found in Gram-positives, and the absence of type III secretion from Gram-positives

Table 16. YscL/FliH Group

Designation[a]	Description	Organism	No. of residues	GenBank GI No.	Accession No.[b]
FliH Bsu	Flagellar assembly protein	*Bacillus subtilis*	208	120329	spP23449
BscL Bbr	Type III secretion protein	*Bordetella bronchiseptica*	212	2935536	gbAF049488
FliH Bbu	Flagellar assembly protein	*Borrelia burgdorferi*	306	1706846	spP52611
YscL Cpn[c]	Type III secretion protein	*Chlamydia pneumoniae*	233	4377138	gbAE001663
YscL Ctr	Type III secretion protein	*Chlamydia trachomatis*	223	3329002	gbAE001327
FliH Eco	Flagellar assembly protein	*Escherichia coli*	228	2506422	spP31068
HrpE Eam	Type III secretion protein	*Erwinia amylovora*	196	1336090	gbU56662
FliH Hpy	Flagellar assembly protein	*Helicobacter pylori*	258	4154848	gbAE001468
PscL Pae	Type III secretion protein	*Pseudomonas aeruginosa*	231	1781394	gbU56077
HrpE Psy	Type III secretion protein	*Pseudomonas syringae (gly)*	193	3603317	gbAF069651
HrpF Rso	Type III secretion protein	*Ralstonia solanacearum*	301	2120693	pirS62087
NolV Rhi	Nodulation protein	*Rhizobium spp.*	208	2498664	spP55716
FliH Sty	Flagellar assembly protein	*Salmonella typhimurium*	235	120330	spP15934
SsaK Sty	Type III secretion protein	*Salmonella typhimurium*	314	3024643	spP74853
FliH Tde[c]	Flagellar assembly protein	*Treponema denticola*	307	2105150	gbU78776
FliH Tpa	Flagellar assembly protein	*Treponema pallidum*	309	3322684	gbAE001218
FliH Vpa	Flagellar assembly protein	*Vibrio parahaemolyticus*	266	4322016	gbAF069392
HrpB5 Xca	Type III secretion protein	*Xanthomonas campestris*	233	984882	gbU33548
YscL Yen	Type III secretion protein	*Yersinia enterocolitica*	223	267573	spQ01253
YscL Ype[c]	Type III secretion protein	*Yersinia pestis*	221	3822053	gbAF074612
YscL Yps[c]	Type III secretion protein	*Yersinia pseudotuberculosis*	221	401694	spQ00928

in turn supports the notion of cooperative function between inner and outer membrane components during the TIIIPS process.

Outer membrane pore-forming secretins serve a variety of functions in addition to functioning as constituents of TIIIPS systems (Marciano *et al.,* 1999). Thus, although they are not constituents of bacterial flagelli, they are constituents of Type II protein secretion (TIIPS) systems (Nouwen *et al.,* 1999), of phage processing systems (Marciano *et al.,* 1999), of pilin (fimbrial) export systems, and of other systems such as the DNA uptake competence system of *H. influenzae*. These secretins generally segregate on their phylogenetic tree according to the type of system with which they interact, suggesting that secretins have not shuffled between the various types of systems for at least the past billion years. However, our phylogenetic analyses show that while the bacteriophage assembly secretins cluster tightly together, and the TIIIPS secretins cluster tightly together, the general secretory (TIIPS) secretins and the fimbrial secretins cluster only very loosely on the secretin phylogenetic tree (Figure12). These facts are consistent with the interpretation that secretins evolved initially as components of the general TIIPS system and as fimbrial export systems, and that they only subsequently assumed functions as constituents of phage and TIIIPS export systems. This suggestion is fully consistent with the expectation that the general secretory system (TIIPS), and maybe fimbrial export systems as well, evolved for essential functions relatively early, and that the TIIIPS and phage assembly systems evolved later employing constituents of the former systems. Our results additionally suggest that intragenic rearrangements during the evolution of multi-domain secretins (Guilvout *et al.,* 1999) has been rare or even non-existent.

These observations and conclusions illustrate the utility of the phylogenetic approach for the purpose of understanding the evolutionary pressures that gave rise to multi-domain, multi-component enzyme transport complexes such as those that comprise the TIIIPS family.

Acknowledgements

We are grateful to Catherine A. Lee for critical and very fruitful discussion and to Mary Beth Hiller and Milda Simonaitis for their assistance in the preparation of this manuscript. Work in the laboratory of MHS was supported by USPHS grants 5RO1 AI21702 from the National Institutes of Allergy and Infectious Diseases and 9RO1 GM55434 from the National Institute of General Medical Sciences, as well as by the M.H. Saier, Sr. Memorial Research Fund.

References

Anderson, D.M., Fout, D.E., Collmer, A., and Schneewind, 0. 1999. Reciprocal secretion of proteins by the bacterial type III machines of plant and animal pathogens suggests universal recognition of mRNA targeting signals. Proc. Natl. Acad. Sci. USA. 96: 12839-12843.

Bonas, U. 1994. *hrp* genes of phytopathogenic bacteria. Curr. Top. Microbiol. Immunol. 192: 79-98.

Buchrieser, C., Brosch, R., Bach, S., Guiyoule, A., and Carniel, E. 1998. The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal asn tRNA genes. Mol. Microbiol. 30: 965-978.

Chen, Y. Smith M.R., Thirumalai, K., and Zychlinsky, A. 1996. A bacterial invasin induces macrophage apoptosis by binding directly to ICE. Embo J. 15: 3853-3860.

Cheng, L.W., and Schneewind, 0. 1999. *Yersinia enterocolitica* type III secretion. On the role of SycE in targeting YopE into HeLa cells. J. Biol. Chem. 274: 22102-22108.

Collmer. A., and Bauer, D.W. 1994. *Erwinia chrysanthemi* and *Pseudomonas syringae* plant puthogens trafficking in extracellular virulence proteins. Curr. Top. Microbiol. Immunol. 192: 43-78.

Feng D.F., and Doolittle, R.F. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. Meth. Enzymol. 183: 375-387.

Franco, A.A., Cheng, R.K., Chung, G.T., Wu, S., Oh, H.B., and Sears, C.L. 1999. Molecular evolution of the pathogenicity island of enterotoxigenic *Bacteroides fragilis* strains. J Bacteriol. 181: 6623-6633.

Frithz-Lindsten, F., Du, Y., Rosqvist, R., and Forsberg, A. 1997. Intracellular targeting of exoenzyme S of *Pseudomoucis aeruginosa* via type III-dependent translocation induces phagocytosis resistance, cytotoxicity and disruption of actin microfilaments. Mol. Microbiol. 25: 1125-1139.

Fu, Y., and Galán, J.F. 1998. The *Salmonella typhimurium* tyrosine phosphatase SptP is translocated into host cells and disrupts the actin cytoskeleton. Mol. Microbiol. 27: 359-68.

Galán, J.E., and Collmer, A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. Science. 284: 1322-1328.

Galyov, E.E., Hakansson, S., Forsberg, A., and Wolf-Watz, H. 1993. A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. Nature. 361: 730-732.

Groisman, F.A., and Ochman, H. 1993. Cognate gene clusters govern invasion of host epithelial cells by *Salmonella typhimurium* and *Shigella flexneri*. EMBO J. 12: 3779-3787.

Groisman, E.A., and Ochman, H. 1997. How *Salmonella* became a pathogen. Trends Microbiol. 5: 343-349.

Guan, K.L., and Dixon, J.E. 1990. Protein tyrosine phosphatase activity of an essential virulence determinant in *Yersinia*. Science. 249: 553-556.

Guilvout, I., Hardie, K.R., Sauvonnet, N., and Pugsley, A.P. 1999. Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? J. Bacteriol. 181: 7212-7220.

Hardt, W.D., Chen L.M., Schuebel, KE., Bustelo, X.R., and Galán, J.E. 1998. S. *typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. Cell. 93: 815-826.

Hermant, D., Menard, R., Arricau, N., Parsot, C., and Popoff, M.Y. 1995. Functional conservation of the *Salmonella* and *Shigella* effectors of entry into epithelial cells. Mol. Microbiol. 17: 781-789.

Hueck, C.J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol. Mol. Biol. Rev. 62: 379-433.

Karaolis D.K., Sornara, S., Maneval, D.R., Jr., Johnson, J.A., and Kaper, J.B. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature. 399: 375-379.

Lee, C.A. 1996. Pathogenicity islands and the evolution of bacterial pathogens. Infect. Agents Dis. 5: 1-7.

Lee, C.A. 1997. Type III secretion systems: machines to deliver bacterial proteins into eukaryotic cells? Trends Microbiol. 5: 148-156.

Lee, V.T., and Schneewind, 0. 1999. Type III machines of pathogenic yersiniae secrete virulence factors into the extracellular milieu. Mol. Microbiol. 31: 1619-1629.

MacNab, R.M. 1999. The bacterial flagellum: reversible rotary propellor and type III export apparatus. J. Bacteriol. 181: 7149-7153.

Marciano D.K., Russel, M., and Simon, S.M. 1999. An aqueous channel for filamentous phage export. Science. 284: 1516-1519.

Mirold, S., Rabsch, W., Rohde, M., Stender, S., Tschäpe, H., Russmann, H., Igwe E., and Hardt, W.D. 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. Proc. Natl. Acad. Sci. USA. 96: 9845-9850.

Mudgett, M.B., and Staskawicz, B.J. 1998. Protein signaling via type III secretion pathways in phytopathogenic bacteria. Curr. Opin. Microbiol. 1: 109-114.

Norris, F.A., Wilson, M.P., Wallis, T.S., Galyov, E.E., and Majerus, P.W. 1998. SopB, a protein required for virulence of *Salmonella dublin,* is an inositol phosphate phosphatase. Proc. Natl. Acad. Sci. USA. 95: 14057-14059.

Nouwen, N. , Ranson, N. , Saibil, H., Wolpensinger, B., Engel A., Ghazi, A., and Pugsley, A.P. 1999. Secretin PulD: association with pilot PulS, structure, and ion-conducting channel formation. Proc. Natl. Acad. Sci. USA. 96: 8173-8177.

Olsen G.J., Woese, C.R., and Overbeek, R. 1994. The winds of evolutionary change: breathing new life into microbiology. J. Bacteriol. 176: 1-6.

Rosqvist, R., Forsberg, A., and Wolf-Watz, H. 1991. Intracellular targeting of the *Yersinia* YopE cytotoxin in mammalian cells induces actin microfilament disruption. Infect. Immun. 59: 4562-4569.

Rosqvist, R., Håkansson, S., Forsberg, Å., and Wolf-Watz, H. 1995. Functional conservation of the secretion and translocation machinery for virulence proteins of yersiniae, salmonellae and shigellae. EMBO J. 14: 4187-4195.

Rossier, 0., Wengelnik, K., Hahn, K., and Bonas, U. 1999. The *Xanthomonas* Hrp type III system secretes proteins from plant and mammalian bacterial

pathogens. Proc. Natl. Acad. Sci. USA. 96: 9368-9373.
Saier, M.H., Jr. 1998. Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. In: Advances in Microbial Physiology. R. K. Poole, ed. Academic Press, San Diego, CA. p.81-136.
Saier, M.H., Jr. 1999. Genome archeology leading to the characterization and classification of transport proteins. Curr. Opin. Microbiol, in press.
Spiik, A.K., Meijer, L.K., Ridderstad, A., and Pettersson, S. 1999. Interference of eukaryotic signalling pathways by the bacteria *Yersinia* outer protein YopJ. Immunol. Lett. 68: 199-203.
Thompson J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25: 4876-4882.
Wachter, C., Beinke, C., Mattes, M., and Schmidt, M.A. 1999. Insertion of EspD into epithelial target cell membranes by infecting enteropathogenic *Escherichia coli*. Mol. Microbiol. 31: 1695-1707.
Young, G.M. Schmiel, D.H., and Miller, V.L. 1999. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. Proc. Natl. Acad. Sci. USA. 96: 6456-6461.