

Research article

Open Access

Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion

Mark J Pallen*, Scott A Beatson and Christopher M Bailey

Address: Bacterial Pathogenesis and Genomics Unit, Division of Immunity and Infection, Medical School, University of Birmingham, Birmingham, B15 2TT, UK

Email: Mark J Pallen* - m.pallen@bham.ac.uk; Scott A Beatson - s.beatson@bham.ac.uk; Christopher M Bailey - cmb036@bham.ac.uk

* Corresponding author

Published: 09 March 2005

Received: 21 December 2004

BMC Microbiology 2005, 5:9 doi:10.1186/1471-2180-5-9

Accepted: 09 March 2005

This article is available from: <http://www.biomedcentral.com/1471-2180/5/9>

© 2005 Pallen et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Like many other pathogens, enterohaemorrhagic and enteropathogenic strains of *Escherichia coli* employ a type-III secretion system to translocate bacterial effector proteins into host cells, where they then disrupt a range of cellular functions. This system is encoded by the locus for enterocyte effacement. Many of the genes within this locus have been assigned names and functions through homology with the better characterised Ysc-Yop system from *Yersinia* spp. However, the functions and homologies of many LEE genes remain obscure.

Results: We have performed a fresh bioinformatics analysis of the LEE. Using PSI-BLAST we have been able to identify several novel homologies between LEE-encoded and Ysc-Yop-associated proteins: Orf2/YscE, Orf5/YscL, rORF8/Escl, SepQ/YscQ, SepL/YopN-TyeA, CesD2/LcrR. In addition, we highlight homology between EspA and flagellin, and report many new homologues of the chaperone CesT.

Conclusion: We conclude that the vast majority of LEE-encoded proteins do indeed possess homologues and that homology data can be used in combination with experimental data to make fresh functional predictions.

Background

Type-III secretion is one of five different types of protein secretion employed by Gram-negative bacteria to move proteins from the cytoplasm across two membranes to the external milieu [1-5]. Any given type-III secretion system (T3SS) consists of a multi-protein complex that spans both the inner and outer membranes and the periplasm so that proteins are delivered to the exterior in an ATPase-dependent fashion without a periplasmic intermediate. Type-III secretion systems can be classified into two major groups: those associated with flagellar biosynthesis and those associated with interactions between bacteria and eukaryotic cells [5]. Type-III secretion is thus central to our

understanding of bacterial motility, virulence, symbiosis, and ecology. Type-III secretion also provides an attractive drug and vaccine target [6] and has been exploited in the biotechnology arena as an antigen delivery system [7,8]

The important human pathogens, enteropathogenic and enterohaemorrhagic *Escherichia coli* (EPEC and EHEC respectively) utilise type-III secretion to subvert eukaryotic signalling pathways by injecting bacterial effector proteins into the host cell cytoplasm [1,9-12]. Within these pathogens, a well-characterised T3SS is responsible for the development of a characteristic attaching-effacing (AE) lesion and for other effects on enterocyte function [9,11-13]

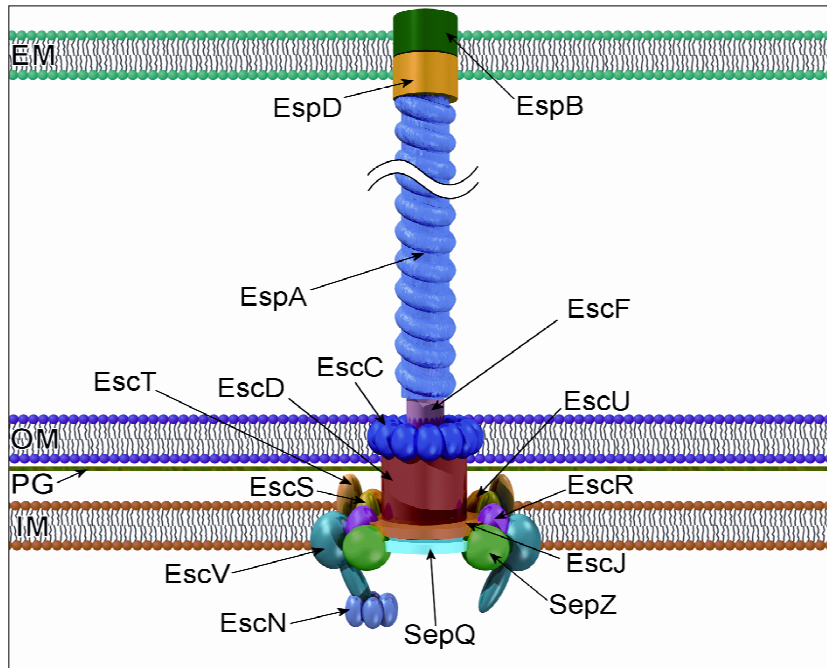
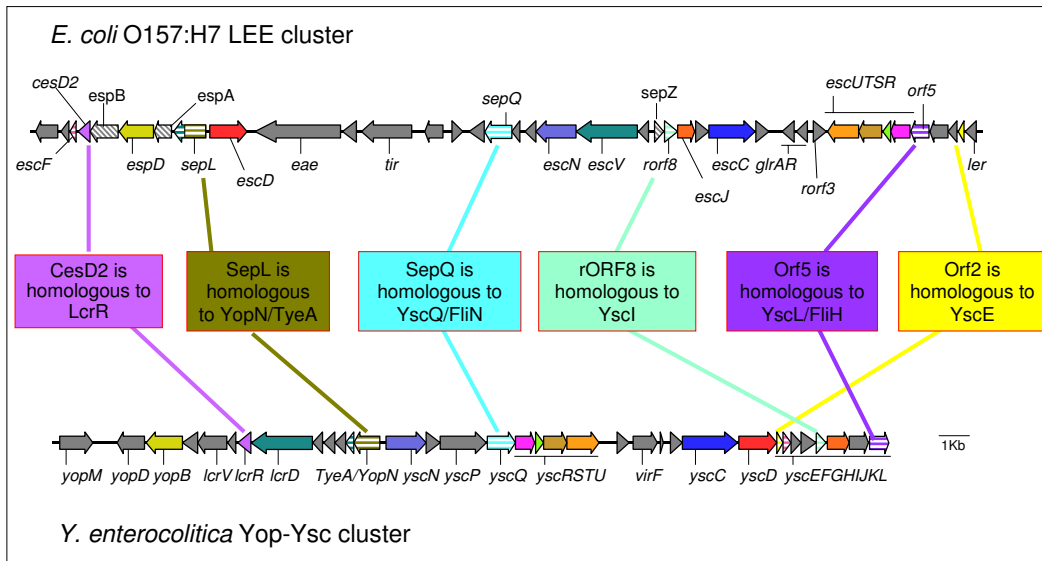


Figure 1

Ia: A comparison between the *E. coli* O157:H7 LEE and the *Yersinia enterocolitica* Yop-Ysc clusters. Ib: Graphical representation of the LEE-encoded type-III secretion system. The LEE region between EspF and Ler is shown (4589000–4620000 of the *E. coli* O157:H7 RIMD 0509952 genome. Accession number: dbj|BA000007.2). The Yop-Ysc cluster between YopM and YscL is shown (15000–40000 of the *Yersinia enterocolitica* plasmid pYVe227 Accession number: dbj|BA000007.2gb|AF102990). Genes shaded in the same solid colour exhibit previously known homologies between the LEE and Ysc/Yop cluster. Genes shaded in the same colour but with horizontal hatching are homologues where we have added information in this paper. Genes which show no apparent homology to any the Ysc/Yop system, but which encode proteins that appear in our graphical representation are shaded with grey/white diagonal hatching (*espB*, *espA*, *sepZ*). Uncharacterized genes where there is no homology, and where their gene products do not appear in our cartoon are shaded solid grey. In the cartoon of the LEE-encoded type-III secretion system proteins are shaded in the same colour as their corresponding genes in the map of the gene cluster. Model based on the type III secretion model from KEGG <http://www.genome.jp/kegg/pathway/map/map03070.html>, and additional published data and images [70, 73, 145, 156, 157]. IM, inner membrane; PG, peptidoglycan layer; OM, outer membrane; EM, eukaryotic membrane.

(Figure 1). In common with most other T3SSs, this system is encoded by a "pathogenicity island" (in this case termed the "locus of enterocyte effacement" or LEE), which contains virulence genes clustered on the chromosome and acquired en bloc by horizontal gene transfer [13-16]. Some strains of the rarely isolated putative pathogen of humans, *Escherichia albertii* (formerly misidentified as *Hafnia alvei*), and of the mouse pathogen *Citrobacter rodentium*, which causes transmissible murine colonic hyperplasia, have been shown to induce AE lesion formation and to possess the LEE [17-20].

When applied to biological sequences, homology is defined as similarity that arises because of descent from a common ancestor [21]. Assignments of sequence homology have a predictive value, in that sequence homology implies structural homology, and, less certainly, functional similarity. Although homologous sequences can diverge in function, or be recruited to a different physiological role or cellular compartment, the discovery of homology allows functional assignments to be transferred from a protein or gene that has undergone experimental investigation to a protein or gene that has not been studied in the laboratory (albeit tentatively and allowing for various common pitfalls in this process [22]). Any such functional assignment should always be treated as a working hypothesis rather than as an established fact. However, this approach can lead to new predictions about biological function that can then be tested in the laboratory, informing an existing programme of experimental work, or even opening up an entire line of enquiry where none existed before [22,23]. At its most basic, an assertion of homology calls to our attention the "null hypothesis" – that proteins with homologous sequences should behave in similar ways – and, whenever this does *not* appear to be true, provokes the question, why are these proteins behaving differently? Searches for homology between sequences are reliable only when the search program provides statistics that allow one to estimate how likely it is that randomly composed sequences could yield alignment scores at least as high as that obtained between the two sequences in question (an example of such a statistical evaluation is the *e* value reported by BLAST [24]).

When the complete sequence of the LEE was first reported, an attempt was made to create a standard rational nomenclature based on experimental findings and functional predictions. *E. coli*-secreted proteins were given the generic name Esp (EspA, EspB, EspD) [11]. Components of the secretion apparatus that had obvious homologues in the better-characterized plasmid-encoded Ysc-Yop T3SS of *Yersinia* sp. were given the prefix Esc and the same suffix as their *Yersinia* homologues (EscC, EscD, EscJ, EscN, EscR, EscS, EscT, EscU and EscV). Proteins that apparently lacked *Yersinia* homologues, but appeared to be involved

in type III secretion, retained the designation Sep (for "secretion of *E. coli* proteins"): originally SepZ for rOrf9 and SepQ for Orf17, but with the designations SepD for rorf6 and SepL for Orf23 subsequently seeing widespread use. The chaperone for the secretion of EspD was named *cesD* ("chaperone for *E. coli* secreted protein D"). The intimin gene was named *eae* (for "*E. coli* attaching and effacing") and the intimin receptor named Tir (for "translocated intimin receptor"). Since these original designations were made, several other LEE-encoded proteins have been characterized and re-named according to these conventions ((EspF, EspG, EspH; CesF, CesAB/CesA, CesD2), while some others have acquired names outside the original nomenclature, based on functional properties (Ler, Map, GrlR, GrlA) [25-33]

The original assertions of homology for the LEE-encoded proteins – and the functional assignments that flowed from them – were based on the results of unsophisticated searches using the gapped BLAST program [11]. However, in the search for distant homologues, a considerable body of evidence now confirms that simple BLAST searches are far outperformed by more advanced, iterative methods such as PSI-BLAST [24,34-39]. In addition, the growth of domain databases such as PFAM [40] and the steady accumulation of new sequence data on type-III secretion systems, especially from genome sequencing [2,23,41], provide a new backdrop against which original claims of homology (or lack of homology) can now be judged. Bearing these facts in mind – and prompted by a recent assertion that "nearly half of the LEE genes have no homologues [33]" – we have therefore undertaken a fresh bioinformatics analysis of the proteins encoded by the LEE, using highly sensitive methods for the detection of homology. Given the recent discoveries of numerous T3SS effectors secreted through the LEE-encoded T3SS [33,42-46], for reasons of space we have opted to restrict our analysis to LEE-encoded components of the secretion and translocation apparatus and soluble cytoplasmic proteins associated therewith (i.e. chaperones and regulators). We will leave homology analyses of effectors (Tir, EspF, EspG, EspH, Map, SepZ [46] encoded within the LEE; potentially many dozens encoded outside the LEE) and functionally related proteins (intimin) for later publications. We conclude that the vast majority of LEE-encoded proteins do indeed possess homologues and that homology data can be used in combination with experimental data to make fresh functional predictions (Figure 1, Table 1).

Results and discussion

The conserved type-III secretion apparatus

Several LEE-encoded proteins have already been assigned Esc designations to reflect their similarities to the Ysc proteins in the *Yersinia* Ysc-Yop system [47]. BLASTP-based comparisons of the predicted components of basal

Table 1: Summary of known characteristics of LEE-encoded genes and their associated proteins, with novel homology-based functional predictions. New names as suggested in body text of this article. Summary of effects on gene expression, type-III secretion and virulence in mice drawn from Deng et al [33]. Yeast two-hybrid results drawn from Creasey et al [145]. Predictions drawn from combination of homology data and cited works. T3S: involved in type-III secretion.

Gene (new name)	Effect of deletion mutation on secretion	Effect on pedestal formation / virulence	Known interactions	Homologues	Domains	Known functions	New functional prediction
<i>rorf1</i>	+	+/+	EspD	Yjik	SdiA-regulated domain (NHL repeats)		SdiA-regulated?
<i>espG</i>	+	+/+		VirA		Translocated effector	
<i>ler</i>	No expression	-/-		H-NS, StpA		Positive regulator	Might form heterodimers with H-NS, StpA
<i>orf2 (escE)</i>	-	-/-	Orf29	YscE, PscE, SsaE, CV2595, YPO0259		T3S	Interacts with C-terminus of SepL; Functionally important conserved residues can be identified by scrutiny of homologues
<i>cesAB</i>	± EspA, EspB	-/-	EspA, EspB	none		Chaperone for EspA and EspB	
<i>orf4</i>	-	-/-		YPO0264, SsaK, CV2589		T3S	Functionally important conserved residues can be identified by scrutiny of homologues
<i>orf5 (escL)</i>	-	-/-		YscL, FliH		T3S	Binds to and regulates activity of ATPase EscN
<i>escR</i>	-	-/-	EscR, EscS, EscU, SepZ, EspD	YscR, FliP	PF00813	T3S	
<i>escS</i>	-	-/-	EscR, EspD	YscS, FliQ	PF01313	T3S	
<i>escT</i>	-	-/-		YscT, FliR	PF01311	T3S	Physically associated in the basal body with EscU in a 1:1 ratio; C-terminus is cytoplasmic [158]
<i>escU</i>	-	-/-	EscrR, EscL, EspD	YscU, FliB	PF00771	T3S	Physically associated in the basal body with EscT in a 1:1 ratio [158]; cleaved at NPTH motif [159]; regulates substrate specificity of secretion system [160].
<i>rorf3 (etgA)</i>	±	±±	EscL	lytic transglycosylase lagB, lpgF			Lytic transglycosylase needed to open gap in peptidoglycan for assembly of secretion system [58]; potential drug target
<i>grlR</i>	+	+/++	GrIR/GrIA	Bongori regulator		Negative regulator	Will work in concert with GrIA
<i>grIA</i>	No expression	-/-	GrIR	Bongori regulator, CaiF		Positive regulator	Will work in concert with GrIR
<i>cesD</i>	-/EspD	±/-		Other TPR chaperones		Chaperone for EspD	TPR chaperone [115]
<i>escC</i>	-	-/-	EscD	YscC, MxiD, InvG		T3S	Interacts with peptidoglycan [161], Dsb-dependent domain missing
<i>sepD</i>	-/Translocators	-/-	SepL	None		Switches translocator/effector secretion	Works in concert with SepL
<i>escJ</i>	-	-/-		YscJ, PrgK, FliF	PF01514	T3S	EscD and EscJ form ring-like structure in inner membrane [66, 162, 163]
<i>rorf8 (escI)</i>	-	-/-	EscU, rORF3	PrgJ, MxiI, YscI		T3S	Forms inner rod within base of needle complex [65, 68-70].
<i>sepZ</i>	+	+±	EscR	None		Translocated effector	
<i>orf12</i>	-	-/-		SsaM		T3S	Might play a role in switching translocator/effector secretion
<i>escN</i>	-	-/-	CesT, Tir [50]	YscN		T3S ATPase	Forms hexameric complex
<i>escV</i>	-	-/-		LcrD, FliA	PF00771	T3S	C-terminal cytoplasmic domain might play a role in gene regulation [164, 165].
<i>orf15</i>	-	-/-		None		T3S	
<i>orf16</i>	±/Translocators	±±		None		Secretion of translocators	FliK/YscP analogue?
<i>sepQ (escQ)</i>	-	-/-		YscQ, FliN	COG1887	T3S	Forms ring within basal body [56]; interacts with EscL [54]
<i>espH</i>	+	+/++				Translocated effector	
<i>cesF</i>	±/EspF?	+/+				Chaperone for EspF	
<i>map</i>	+	+/++	CesT			Translocated effector	
<i>tir</i>	+	-/-	CesT Tir			Translocated effector	
<i>cesT</i>	±/Tir	-/-	Map, EspF, Tir, CesT	Many		Chaperone for Tir	Possible regulatory role, by analogy with ExsC?
<i>eae</i>	+	-/-				Adhesin (intimin)	

Table 1: Summary of known characteristics of LEE-encoded genes and their associated proteins, with novel homology-based functional predictions. New names as suggested in body text of this article. Summary of effects on gene expression, type-III secretion and virulence in mice drawn from Deng et al [33]. Yeast two-hybrid results drawn from Creasey et al [145]. Predictions drawn from combination of homology data and cited works. T3S: involved in type-III secretion. (Continued)

<i>escD</i>	-	-/-		YscD, PrgH	YscD: FHA and BON domains	T3S	EscD and EscJ form ring-like structure in inner membrane [66, 162, 163]; BON domains mediate binding to phospholipids; FHA domain mediates protein-protein interactions and signalling
<i>sepL</i>	-/Translocators	-/-	SepD	YopN, TyeA		Switches translocator/effector secretion	single protein mediating effects of YopN and TyeA; internal interaction between YopN- and TyeA-like moieties; also a translocated effector?
<i>espA</i>	+	-/-	CesAB	flagellin		Translocator	Half-flagellin model (see text); dimerizes before polymerizes; lacks D0 and D3 domains
<i>espD</i>	+	-/-	EspD, EscRSU, CesD, rORF1, SepZ	YopB etc		Translocator	
<i>espB</i>	+	-/-	CesAB	YopD etc		Translocator	
<i>cesD2</i>	±/EspD?	+/±		LcrR		Chaperone for EspD	LcrR will bind YopB
<i>escF</i>	-	-/-		YscF		Major needle component	
<i>orf29</i>	-	-/-	Orf2	YPO0261, Ssal, H, CV2586		T3S	Functionally important conserved residues can be identified by scrutiny of homologues
<i>espF</i>	+	+/+				Translocated effector	

complex confirms significant near-full-length homology between EscC, EscD, EscJ, EscN, EscR, EscS, EscT, EscU and EscV and the respective Ysc proteins and homologues (Figure 1, Table 1), although, confusingly, YscV is more commonly known as LcrD, and EscD is sometimes termed Pas [1,47,48]. In the cases of EscN, EscR, EscS, EscT, EscU and EscV, homology clearly extends to components of the flagellar apparatus [1]. These unequivocal homologues suggest that observations on these proteins in other systems can safely be generalised to the LEE-encoded system – as has been confirmed in several recent papers [49,50] – and can be used to frame novel hypotheses about their function (Table 1, Figure 1). However, it is important to remain on guard for minor but, perhaps, important differences between systems: for example, the C-terminal ~100-residue domain from YscC, which is thought to house two DsbA-dependent disulfide bridges is missing from EscC and many other homologues [51].

As noted previously, YscD contains a cytoplasmic FHA domain [52]. PSI-BLAST searches confirm that EscD also contains a cytoplasmic FHA domain (Figure 2). In addition, these searches revealed the presence of at least one putative phospholipid-binding domain (also called a BON domain [53]) in the periplasmic portion of the protein (Figure 3). Two compelling hypotheses arise from these observations. The first of these is that, given the established role of FHA domains in signalling [52], it is tempting to speculate that the cytoplasmic FHA domain of EscD might be involved in signal transduction (alternatively, it might represent a molecular fossil, now fully adapted to a purely structural role). Secondly, as BON domains are thought to mediate binding to phospholip-

ids in a variety of other proteins [53], it is likely that they play a similar role in EscD and its homologues.

The missing EscL and EscQ proteins

Although, as noted above, the majority of the "Ysc proteins" that form conserved components of the secretion apparatus in Yersinia have easily recognisable "Esc protein" counterparts in the LEE-encoded system, there are two salient exceptions – in the current annotation of the LEE, there are no equivalents of YscQ and YscL. YscQ is a member of the FliN family of proteins and is thought to be a component of the basal secretion complex [54]. Although PSI-BLAST searches fail to find any YscQ homologue among the LEE-encoded proteins, a search of the NCBI's CDD database [55] identifies a FliN domain (COG1886) in the C-terminal half of SepQ (Figure 4). This suggests that SepQ plays a similar role to YscQ and FliN – allowing new hypotheses to be framed (Table 1) – and should be re-named EscQ. Although, the high degree of divergence of the EscQ C-terminal domain from other YscQ/FliN family members is puzzling, it is clear from scrutiny of the recently solved structure of the homologous domain from HrcQB [56] that SepQ possesses most of the conserved motifs common to this domain family, and must adopt a similar fold (Figure 4).

YscL is a member of the FliH family of proteins. In flagellar systems, FliH binds to and regulates the activity of the ATPase FliI [57], and YscL is also known to bind to YscN [54]. PSI-BLAST searches across the NCBI's non-redundant database with YscL fail to identify any homologue in the LEE system, because of contamination with low-complexity eukaryotic proteins. However, if a PSI-BLAST search with Orf5 is restricted to bacterial proteins, after

EHEC EscD	26	-LTLGTED---NDIVYFPLE---QGLNQFLLD---IREEG---VFLSPV-----EFWIDGQPT	71
Vp_PscD	25	-LVLGSDD---FIADLVLS DAG---VEANHFTLV---CDSES---VMLRGCQ-----DVTINGENR	72
Pa_PscD	24	-VALGSDP---LQADLVLLDEG---IAAVHLVLE---VDAQG---VRLLEWA---EGCEPRQDGOAQ	74
Ah_AscD	24	-LVIGSDP---LKADLVLVDEG---MAAEHLVLL---VSTEG---ITLQAWA---DGITPTQDGVVAL	74
Pl_LscD	24	-LAIGSDP---LQADIVLVDEDED---VAPAHVLE---VEPDS---IRLLEWA---GETAPSQNGEPL	74
Ye_YscD	24	-CVFGSDP---LQSDIVLSDSE---IAPVHLVLM---VDEEG---IRLTD SA-----EP LLOEGLPV	81
Dv_YscD	28	-HVI GTED---SCDIILSDGG---LAPRHVSI G---VTAGV (6) VHTPLD-----GPVRYDDAEV	75
Ct_CT644	25	-WSLGS DG---KASDILLQDEK---LAPSQIRIT---LKDGE---YYLENLD---ALRPVSV DGTVI	75
Cp_CpB0739	25	-WSIGRDS---SANDIPIEDPK---LGASQAIIN---KTDGS---YYITNLD---DTIPIVNGVAI	70
Ea_HrpQ	23	-WWIGAAQ---DADLALFDPG---IKDRHCRLS---KTDLG---WEVTALE-----GPLNDNEGQR	70
Xa_HrpD5	24	-YTLGSDP---QTDIRIDDW---PGCSLIE---VDQDQ---IRYSSDT-----VPETSFVAFQ	70
AF6_HUMAN	1	VTEVGT EKLD---DNSIQ LFGPG---IQPHCDLT---NMDGV---VTVTPRSM---DAETYVEGQRI	53
S74610	1	TYSIGRES---SNDI VIYDQV---VSRHHATLI---RIKPT (7) YRLDGD (5) STNGLIINGQDS	63
YD6A_SCHPO	1	PIYIGRYT (6) DVSAIVFRSKV---VSRRH AQIF---YENN---TWYIQDMG---SSSGTFLNHVRL	58
CEC44B9	1	VTKLGT AH---DCHVRC DKLT (7) LADYHCDIV---RDYYS---DNFYISTL---GTEAVFVDGRAV	59
MEK1_YEAST	1	LVKVGRNDK---ECQLVLTNPS---ISSVHCVFVCFDEDSIPMFYVKDCS---LNGTYLNGLL	57
I46033	1	YYLFG RNP D---LCDFTIDHQS---CSRVHAALV---YHKHL---KRFLIDLN---STHGTFLGHIRL	54
DMKINESIN	1	LVRIGRGR (8) QPDIVLDGPL---VALQHCSI EHE---RGGK---LYVIPGS---EDFETYVNGELL	60
SPK1_YEAST	2	PFFIGRSE---DCNCKIEDNR---LSRVHCFIF (4) AVGKS (5) AQLDDI (5) GTNVSYLN NNRM	66
A61188	1	TYRLGHRAD---LCDVALRPQQ (4) ISGIHAELHAE---PRGDD---WRVSL EDHS---SQGTLVNNVRL	59
YOT2_CAEEL	1	FVVI GR IKP---GCDLLMEHPS---ISRYHCILQ---YGNDK (7) WHIFELG---STHGSRMNKKRL	59
AF015309	1	EITLGRIT---KDNQIDVDLSL (6) ISRKQGVIK---LKNNGD---FFIANEG---RRPIYIDGRPV	58
YKI5_CAEEL	1	GKSFGRAT (9) TEHVELEPQF---ISRCHARVHH---TNQDGVVEEYLVEDIS---ENGTYINDRRL	63
Consensus/70%		.h.l.tpps....scl.l.s.....lt..+h.l.....p..s...hbl.shs.....shhss..b	

Figure 2

Multiple alignment of the FHA domain from EscD with other FHA domains. Alignment is presented using default CHROMA [155] settings: Consensus abbreviations (amino acids): a, aromatic (FHWHY, blue lettering on a dark yellow background); b, big (EFHIKLMQRWY, blue on light yellow); h, hydrophobic (ACFGHILMTVWY, black on dark yellow); l, aliphatic (ILV, grey on dark yellow); p, polar (CDEHKNQRST, blue on white); s, small (ACDGNPSTV, dark green on white); t, tiny (AGS, light green on white); -, negatively-charged (DE, red on white); and, +, positively-charged (KR, blue on white), c, charged (DEKRH, pink on white). Organism and gene name abbreviations used: EHEC (*Escherichia coli* O157:H7) EscD (ECs4558/dbj|BAB37981), Vp (*Vibrio parahaemolyticus*) PscD(ref|NP_798074), Pa (*Pseudomonas aeruginosa*) PscD (ref|NP_250408), Ah (*Aeromonas hydrophila*) AscD(gb|AAS91829), Pl (*Photobacterium luminescens*) LscD(gb|AAO18032), Yp (*Yersinia enterocolitica*) YscD(gb|AAC37021), Dv (*Desulfovibrio vulgaris*) YscD(ref|YP_009153), Ct (*Chlamydia trachomatis*) CT664 (NP_220183), Cp (*Chlamydomonas reinhardtii*) CpB0739(gb|AAP98668), Xa (*Xanthomonas axonopodis*) HrpD5 (gb|BAD29996), Ea (*Erwinia amylovora*) HrpQ(gb|AAB06000). The remaining 12 sequences are representative members of the SMART FHA domain (SM00240). Gene names, numbering and alignment are as presented in the SMART FHA family alignment <http://www.smart.ox.ac.uk/>.

one iteration, YscL appears in the results list (15% identity 29/184 residues; e value 0.002). After two iterations, several more FliH homologues are found. Furthermore, consistent with the recent suggestion that the YscL-YscN interaction mirrors similar interactions in other ATPases [54], weak similarity is also reported between Orf5 and several F⁰ ATPase beta subunits (data not shown). A multiple alignment confirms the presence of conserved residues within FliH, YscL and Orf5 (Figure 5). It thus seems likely that Orf5 is a homologue of YscL and FliH, plays a similar role (Table 1, Figure 1) and should be re-named EscL.

A lytic trans-glycosylase: a novel drug target?

Both domain searches and PSI-BLAST searches show that rOrf3 encodes a lytic transglycosylase. Similar enzymes,

which occur in type II, type III and type IV secretion systems, are responsible for enlarging gaps in the peptidoglycan meshwork to allow efficient assembly and anchoring of supramolecular transport complexes in the cell envelope [58]. The presence of a dedicated transglycosylase within this type-III secretion system provides an obvious target for the development of anti-microbial agents that specifically inhibit type-III secretion, particularly as structural data is available for homologous enzymes [59-61]. We propose that rOrf3 be re-named EtgA for *E. coli* transglycosylase. Curiously, we have been unable to find a homologue of EtgA in the Ysc-Yop system.

The needle

Despite its name, EscF, a major component of the needle, does not show significant similarity to YscF on a simple

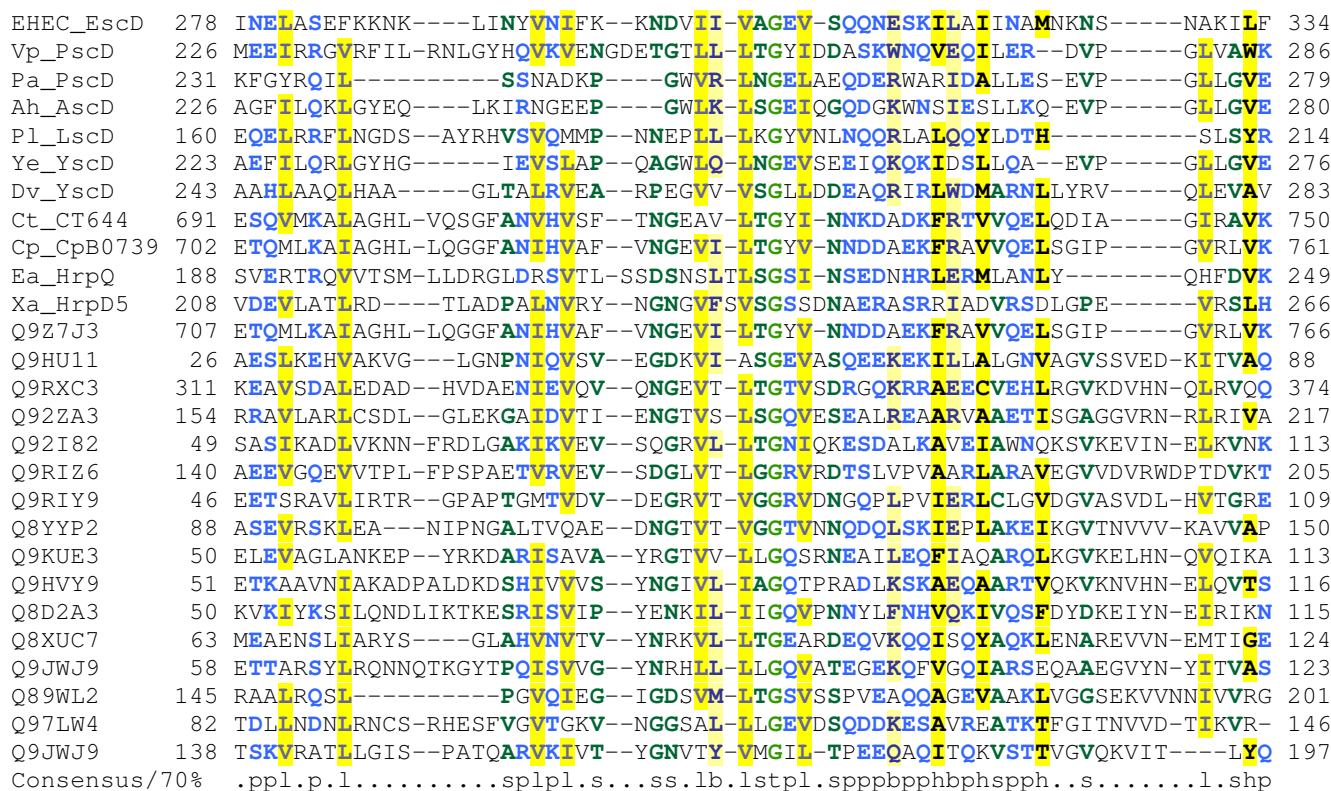


Figure 3
A multiple alignment of the phospholipid-binding (BON) domain from EscD with other BON domains. The most obvious BON domain from EscD is shown. However, from patterns of residue conservation centred on conserved glycines in the EscD family of proteins (data not shown), we suspect that there may be one or two more BON-like domains within the cytoplasmic portion of EscD. For sequences 1–11 organism and gene name abbreviations are as for Figure 2 legend. The remaining 16 sequences are representative members of the Pfam BON domain (PF04972). Gene names, numbering and alignment are as presented in the Pfam:BON family alignment <http://pfam.wustl.edu/>. Alignment is presented using default CHROMA settings (see Figure 2 legend).

BLASTP search (19% identity; e value of only 0.88), but the unimpressive significance score can be explained by the short length of the sequence (only 73 residues); the e value rises to a more respectable 0.003 within two iterations of a PSI-BLAST search. Furthermore, it is clear from published experimental work that YscF and EscF play equivalent roles as needle components [62-64].

In the Spi-1 and Mxi-Spa systems, two small proteins are associated with the needle – PrgI/PrgJ and MxiH/MxiI respectively [65-68]. There is experimental evidence from both systems to suggest that only one of the proteins (PrgI/MxiH) is the major subunit component of the needle, while the function of the second protein (PrgJ/MxiI) was, until recently, unclear – it was initially thought to form a cap for the needle [65,68,69]. However, the function of PrgJ (and by extrapolation, of its homologues) has

now been elucidated – it forms the inner rod within the base of the needle [70].

A PrgJ-like component has yet to be described in the LEE-encoded system. However, PSI-BLAST searches indicate that rORF8 is homologous to EprJ, PrgJ, MxiI and to an uncharacterised protein from the Ysc-Yop system, YscI (Figure 6). This suggests that rORF8 and YscI play similar roles to PrgJ/MxiI as components of the inner rod and that rORF8 should be re-named EscI. Interestingly, PSI-BLAST searches also suggest that the PrgI-like and PrgJ-like proteins are paralogous.

Homology between EspA and Flagellin

The LEE-encoded T3SS is so far unique in possessing a filamentous organelle, the EspA pilus, that resembles the flagellum, forms an extension to the needle complex and is thought to act as a molecular syringe providing a

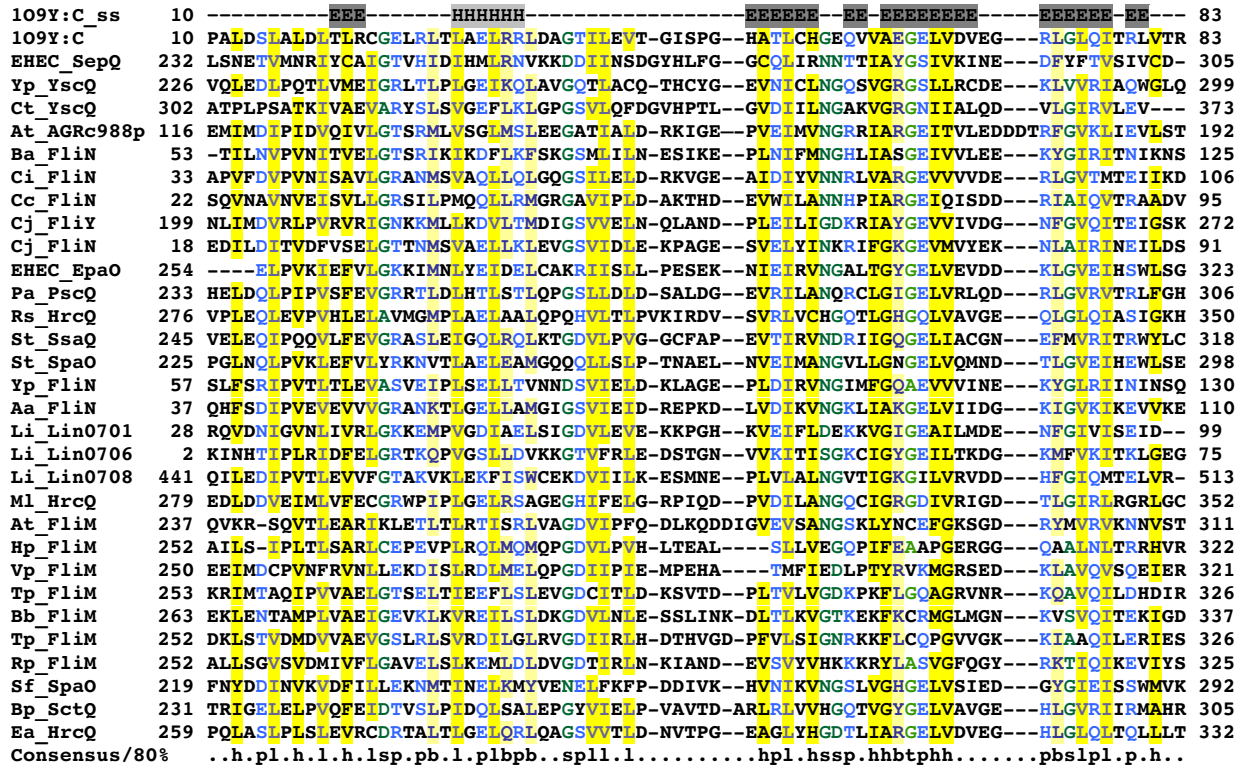


Figure 4
Multiple alignment of SepQ/EscQ with related type-III secretion proteins. The sequence and secondary structure of the crystal structure of the C-terminal domain of the HrcQb protein from *Pseudomonas Syringae* Pv. *Phaseolicola* (pdb|1OY9) are shown 56. Sequences 2–21 (including SepQ) are from COG1886. 1OY9 and the remaining 10 sequences are representative members of Pfam:SpaO. Organism and gene name abbreviations are abbreviated as follows: EHEC (*Escherichia coli* O157:H7) SepQ (ECs4565/gb|AAG58829), Yp (*Yersinia pestis* CO92) YscQ (ref|NP_403921), Ct YscQ (ref|NP_220191), At (*Agrobacterium tumefaciens* str. C58) AGRc988p(ref|NP_353589), Ba (*Buchnera aphidicola*) FliN (sp|P57183), Ci (*Caulobacter vibrioides*) FliN(sp|Q03593), Cc (*Caulobacter crescentus*) FliN(ref|NP_420978), Cj (*Campylobacter jejuni*) FliY (ref|NP_281274) and FliN (ref|NP_281542), EHEC EpaO (ECs3726/sp|Q8X6F0), Pa (*Pseudomonas aeruginosa*) PscQ (ref|NP_250385), Rs (*Ralstonia solanacearum*) HrcQ (ref|NP_522422), St (*Salmonella typhimurium*) SsaQ(sp|P74860) and SpaO (sp|P40699), Yp FliN(ref|NP_404342.1), Aa (*Aquifex aeolicus*) FliN (sp|O67495), Li (*Listeria innocua*) Lin0701 (ref|NP_470044), Lin0706 (ref|NP_470049), and Lin0708 (ref|NP_470051), Ml (*Mesorhizobium loti*) hrcQ (ref|NP_106868), At FliM (sp|Q44457), Hp (*Helicobacter pylori*) FliM (sp|O25675), Bs (*Bacillus subtilis*) FliM (sp|P23453), Bb (*Borrelia burgdorferi*) FliM(sp|57511), Tp (*Treponema pallidum*) FliM (sp|P74927), Vp FliM (sp|Q9Z6GI), Rp (*Rhodobacter sphaeroides*) FliM (O85118), Sf (*Shigella flexneri*) SpaO (sp|P35534), Bp (*Burkholderia pseudomallei*) SctQ (Q9ZGR1), Ea (*Erwinia amylovora*) HrcQ (Q46645). Alignment is presented using default CHROMA settings (see Figure 2 legend).

continuous protein-translocation channel between the bacterial and host cells [71-73]. The pilus primarily consists of EspA, but also requires EspD for its biosynthesis [72]. EspB and EspD form a pore in the host cell plasma membrane [74]. PSI-BLAST searches on the ViruloGenome site with EspA reveal homology within two iterations to numerous flagellins, including the *Salmonella* flagellin for which a structure is available [75,76] (Figure

7). Although the precise alignment of the N-terminal region of EspA with flagellin appears critically dependent on the PSI-BLAST settings used, the alignment of the C-terminal region of EspA with flagellin appears robust. This allows certain hypotheses to be framed about the structure of EspA and the EspA filament.

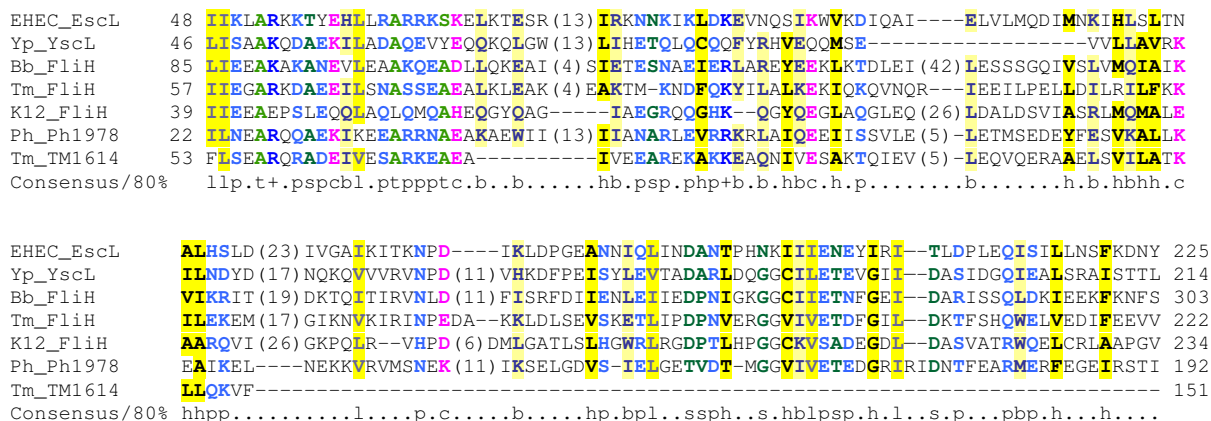


Figure 5
Multiple alignment of Orf5/EscL with related type-III secretion proteins. Organism and gene name abbreviations: EHEC (*Escherichia coli* O157:H7) EscL (ECs4584/sp|O85643), Yp (*Yersinia pestis* CO92) YscL(ref|NP_395195), Bb (*Borrelia burgdorferi*) FliH(ref|NP_212423), Tm (*Thermotoga maritima*) FliH(ref|NP_228034), Ph (*Pyrococcus horikoshii* OT3) Ph1978(ref|NP_143803), *E. coli* K-12 FliH (ref|NP_416450), Tm TM1614(ref|NP_229414). Note that Ph1978 and TM1614 are annotated as a hypothetical protein and an ATPase synthase F0, subunit b, respectively. Alignment is presented using default CHROMA settings (see Figure 2 legend).

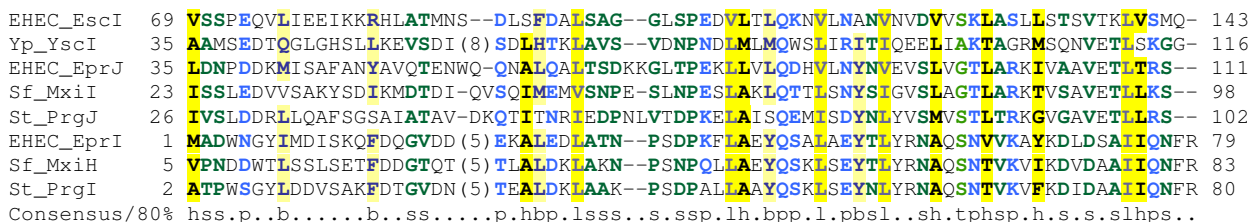


Figure 6
Multiple alignment of rOrf8/EscI with related type-III secretion proteins. Organism and gene name abbreviations as follows: EHEC (*Escherichia coli* O157:H7) EscI (ECs4572/sp|O85634) and EprI (ECs3717/sp|Q8X6G5), Sf (*Shigella flexneri*) MxiI (sp|Q06080) and MxiH (sp|Q06079), St (*Salmonella typhimurium*) PrgJ (sp|P41785) and PrgI (sp|P41784), and Yp (*Yersinia pestis* CO92) YscL (Q00933). Alignment is presented using default CHROMA settings (see Figure 2 legend).

Firstly, surprisingly, the coiled-coil domain at the C-terminus of EspA appears to correspond *not* to the coiled-coil at the extreme C-terminus of flagellin, which forms one half of the flagellin D0 domain, but *instead* to the C-terminal portion of the D1 domain (CD1) [75,76]. In other words, EspA appears to lack a D0 domain. In the flagellar filament, D0 forms the inner tube, while D1 forms the outer

tube [75]. Although D0 interactions are important in filament stability, mutant flagellins that lack the D0 domain can still assemble into straight filaments, albeit with a structure, termed Lt, distinct from the native flagellar filaments [77]. It is thus entirely plausible that the EspA pilus is assembled without the need for a D0 domain and resembles the Lt filament structure. One problem with

A third, crucial point concerns the role of the coiled-coil domains in EspA. The coiled-coil domains at the N- and C-termini of flagellin were initially thought to facilitate filament assembly by mediating inter-molecular interactions between neighbouring flagellin subunits [79,80]. Based on the hypothesis that coiled coils might play an analogous role in the assembly of the EspA pilus [72,81], mutagenesis experiments were performed and showed that the C-terminal coiled coil domain of EspA is required for filament assembly [82]. However, the interpretation of these experiments must now be revised in the light of the flagellar structure and homology data. When the flagellar filament structure was solved, it became clear that, contrary to expectation, the coiled-coil domains mediated interactions *within* subunits, rather than *between* them [76]. Thus, by extension, it is likely that, for both EspA and flagellin, the coiled-coil interaction is a necessary but not sufficient requirement for filament assembly.

Coiled-coil prediction programs and PSI-BLAST searches suggest that, in addition to lacking the D0 and D3 domains, EspA lacks any homologue of the N-terminal helix (ND1) of the coiled-coil that forms the D1 domain. However, recent structural information on EspA complexed with its chaperone CsaA indicates that EspA contains a short N-terminal coiled-coil domain in addition to the already recognized much longer C-terminal coiled-coil domain [83]. It is thus possible that EspA folds up as a "mini-flagellin", so that together these two coiled-coil domains mediate an intra-subunit interaction similar to that seen in the flagellar monomer. However, this does not fit well with the apparent discrepancy in length between the two domains. An alternative possibility is that EspA acts as "half-a-flagellin", so that it dimerizes through an inter-subunit coiled-coil interaction, and then the dimer polymerizes, like flagellin, through hydrophobic interactions but between dimers rather than monomers.

The "mini-flagellin" and "half-flagellin" models could be distinguished in the laboratory. For example if the half-flagellin model holds true, it should be possible through mutagenesis to identify mutations that disrupt polymerisation but leave the coiled-coil-mediated dimerization potential intact (an approach that might make it possible to grow crystals for structure determination). Furthermore, it should be possible to distinguish dimerization from subsequent polymerisation by physicochemical methods. Finally, under this model, there is a potential for the formation of heterodimers – an idea supported by the existence of four *C. violaceum* EspA homologues, encoded by two pairs of adjacent genes. In each pair, the first gene encodes a short conventional EspA homologue, while the second gene encodes a long EspA homologue containing a central D3-like insertion [41]. According to the half-flag-

ellin model, the short and the long EspA-like proteins would have the potential to form heterodimers, which would then polymerise into a filament. This prediction could be tested experimentally, e.g. by showing that these EspA homologues form long/short dimers but do not engage in self-self interactions.

The Toll-like receptor 5 (TLR5) recognizes the D1 domain of flagellin [84]. Although EspA lacks several of the residues identified as critical in binding (data not shown), the overall conservation of the CD1 domain raises the question of whether TLR5 might recognize EspA. However, recent experimental investigations suggest it does not (K. D. Smith, personal communication). Yet, this in turn raises the question of whether this lack of interaction is the result of genetic drift or of positive selective pressure on EspA to avoid recognition by TLR5.

PSI-BLAST searches with EspB prove unhelpful, as the compositional bias of the protein, particularly in its coiled-coil domains, attracts numerous simple-sequence proteins, even when a filter and composition-based statistics are used [81,85]. However, a comparison of domain architecture and genetic location with translocation-pore proteins/genes from other systems (e.g. SipC, IpaB, YopD) suggests functional and structural similarities. PSI-BLAST searches with EspD prove more fruitful, especially when composition-based statistics are employed – they reveal significant similarities to several other T3SS translocon-pore proteins. For example, when EspD is used as a query sequence for a PSI-BLAST search on the NCBI NR database under default settings, YopB is found in the first round with an *e* value (0.001) within the PSI-BLAST threshold, while SseC achieves significance (4e-07) on the second PSI-BLAST round.

Homologies and Gene fusions in the *SepL/YopN-TyeA* family

SepL is a LEE-encoded protein that has been said to have no homologues and is required for AE lesion formation and secretion of translocation proteins and for the translocation of – but not the secretion of – effectors [86,87]. PSI-BLAST searches reveal homology to two proteins from the Ysc-Yop system: most of the SepL sequence (up to residue 267) is homologous to YopN (sometimes called LcrE), while the C-terminal 83 amino acids are homologous to TyeA (Figure 8). This assertion of homology between SepL and YopN/TyeA is supported by several other lines of evidence (i) the *yopN* and *tyeA* genes are adjacent to one another; (ii) the YopN and TyeA proteins interact with one another; (iii) it is well established that genes encoding interacting domains often occur adjacent to one another and can undergo gene fusion in some genomes, (iv) experimental evidence has recently emerged showing that a chimaeric YopN/TyeA protein

can be produced by programmed frame shifting in some *Yersinia* strains [88-91].

This new-found homology also fuels two predictions: (i) the C-terminal residues 268–351 of SepL, corresponding to TyeA, interact with the residues immediately upstream (229–267) in the region of SepL corresponding to YopN residues 242–280; (ii) the interaction between YopN and TyeA must occur in a 1:1 ratio – this could be confirmed by creating a *yopN-tyeA* fusion that lacks a start codon for *tyeA*, and then showing that this construct could complement a *yopN/tyeA* mutant to give normal type-III secretion.

Further insights into type-III secretion arise when additional PSI-BLAST searches are applied to the homologues retrieved when SepL is used as the query sequence. Firstly, it is clear that the Ysc-Yop system is unusual in separating out the two SepL-related sequences into two proteins. This arrangement occurs only in systems closely related to the Ysc-Yop system (i.e. T3SSs from *Pseudomonas aeruginosa*, *Photobacterium luminescens*, *Aeromonas salmonicida*, *Vibrio parahaemolyticus*); in other systems, there is a single protein corresponding to a YopN-TyeA chimera (e.g. SsaL in Spi-2, HrpJ in *Erwinia* and related species, BopN in *Bordetella*, CopN in *Chlamydia* [92]). Secondly, through careful use of PSI-BLAST, it is possible, through repeated searching, to establish unequivocal homology between SepL-like proteins from each of the well-characterised animal-pathogen-related T3SSs – not just between SepL and YopN-TyeA, but also encompassing InvE from the Spi-1 system and MxiC from the Shigella system (Figure 8). This implies that functional assignments for any one of these proteins should be generalisable to all the others – a prediction borne out by the similar phenotypes ascribed to these proteins, which all appear to play a role in regulating the translocation of effectors [86,87,93-98]. Indeed, it is even possible that some reported differences in phenotype (e.g. the surface location of YopN versus the apparent cytoplasmic location of InvE; the fact that YopN appears to control secretion of all Yops, whereas InvE is more selective in controlling secretion of translocators rather than effectors [98]) might represent idiosyncrasies of experimental approach than genuine biological differences. However, even if these do represent true differences, this begs the question: how and why have homologous proteins come to acquire such divergent functions?

SepL interacts with another protein of unknown function SepD [87]. Unfortunately, homology searches afforded no insights into the function of SepD. However, given that *sepD* is located next to *escC* in the LEE, while *yscB* occupies an analogous position (next to *yscC*) in the *ysc* cluster, it is tempting to speculate that they might play similar roles.

Also, no homologues could be found in the LEE-encoded system of LcrV and LcrG, which are thought to regulate entry into the Ysc secretion apparatus from the cytoplasm [99,100]. Similarly, we could not detect any LEE-encoded homologues of the needle-length molecular ruler YscP [101], even though, as we have reported elsewhere [23], PSI-BLAST searches do reveal sequence similarity at the C-terminus between YscP and the flagellar hook-length-determining protein FliK, implying that they employ similar mechanisms [102,103]. Nor could we find any homologue of InvJ/Spa32, which, although showing no detectable sequence similarity to YscP or FliK, is thought to play a similar role to these proteins in needle-length determination in the Spi-1 and Mxi-Spa systems [67,104].

Despite the frustrating lack of sequence similarities, two pieces of tantalising circumstantial evidence link FliK/YscP with InvJ/Spa32, and hint at the identity of an equivalent protein in the LEE-encoded system. Firstly, all four proteins are characterised by the presence of large amounts of random coil on secondary structure predictions (S. Lloyd, personal communication). Secondly, the genes for these proteins always occupy an equivalent position within the type-III secretion gene cluster: always next door but one to and downstream of the gene encoding the ATPase. When these criteria are applied to the LEE, an obvious candidate for needle-length determination emerges: Orf16 (28% random coil according to SOPMA [105]). Almost nothing is known about this gene/protein, which has no known homologues outside of the LEE. However, a mutant in this gene in *Citrobacter* is deficient in translocator secretion, but is still able to secrete effectors normally, and is severely attenuated in the mouse [33]. These results are consistent with the hypothesis that Orf16 controls the switch between export of needle components and components of translocation apparatus. However, in the absence of structural or experimental data, it is premature to make an assertion of homology between Orf16 and YscP/FliK, particularly as tentative assignments tend to harden into dogma and then preclude consideration of other possibilities i.e. that different systems employ different mechanisms for the same purpose. Clearly, additional experimental data are needed to resolve this issue.

Chaperones

Five LEE-encoded proteins have been designated chaperones on the basis of experimental work: CesT, CesF, CesD, CesD2 and CesAB [29-31], [106-108]. For one of these, CesT, which chaperones the effectors Tir and Map, the structure is available [109]. PSI-BLAST searches with the CesT sequence link it to CesF and to two other virulence-related T3SS chaperones for which structural information is available (SycE, SicP, SigE [109-112]) and also identify numerous other uncharacterised T3SS-associated proteins

as chaperones (Figure 9). Interestingly, the results show homology between CesT/CesF and ExsC from *P. aeruginosa*. As this protein has recently been reported to act as an anti-anti-activator, controlling transcription of T3SS genes in *P. aeruginosa* [113,114], it is plausible that CesT and/or CesF might play similar roles in gene regulation.

CesD is a member of the "chaperones of the translocators" family common to many T3SSs. We have already shown that this family is characterised by the presence of three tandem tetratricopeptide repeats [115]. CesAB (previously Orf3) is a newly described chaperone that binds to EspA and EspB [31]. Homology searches fail to reveal any homologues of this protein, even in the two other systems that possess obvious EspA homologues.

CesD2 (previously Orf27), is a recently described chaperone that, like CesD, binds to EspD [29]. PSI-BLAST searches show that it is homologous to SseE from the Spi-2 system, CseE from the Cpi-2 system of *Chromobacterium violaceum*, LcrR from the Ysc-Yop system, AcrR from *Aeromonas salmonicida* and LssR from *Photobacterium luminescens*. None of these homologues has been characterised aside from LcrR. An initial report suggested that a mutation in *lcrR* had a regulatory effect [116], although this was subsequently shown to be the result of a polar effect on the downstream gene *lcrG* [117]. However, an *lcrR* mutant was detected in a signature-tagged mutagenesis screen of *Y. pseudotuberculosis* and shown to be attenuated both after oral and intra-peritoneal infection in Balb-C mice, even though the mutation did not have a measurable effect on Yop secretion or function in cell culture assays [118]. These findings stand in contrast to a report that a *Y. pseudotuberculosis lcrR* mutant had no virulence phenotype after oral infection of Swiss albino mice [119]. The homology between LcrR and CesD2 and between EspD and YopB suggests that LcrR is likely to bind to YopB (Table 1).

Regulators

Ler is a regulatory protein encoded with the LEE1 operon, which has been extensively studied for its effects on gene expression within the LEE-encoded system [120-123]. It is known to be homologous to the global regulator H-NS, and is likely to adopt a similar fold, including a dimeric structure, particularly as key residues in the N-terminal dimerization domain are conserved in both proteins [124-126]. H-NS is known to form heterodimers with other proteins from the H-NS family [124,127,128] and to form complexes with members of the Hha family of regulators [129]. As the attaching and effacing pathogens possess two other H-NS-like proteins besides Ler (H-NS and StpA), and possess Hha-like proteins (Hha, YdgT), it is tempting to ask whether Ler could, under any circum-

stances, form complexes with H-NS or StpA or with the Hha-like proteins (and if not, why not?).

Quorum sensing through the AI-2 pathway is known to influence type-III secretion in the LEE-encoded system [130-134]. In addition, there is some limited evidence to support an impact on type-III secretion of quorum-sensing through the AI-1/acyl-homoserine lactone route [135]. However, this latter route is less plausible, as *E. coli* cannot produce these signalling molecules, although it can probably sense signals produced by other bacteria through the LuxR-like protein SdiA [136]. It is thus surprising that domain searches with the protein sequence from rOrf1 of the LEE identify a PFAM domain (PF06977) entitled "SdiA-regulated", which represents a conserved region approximately 100 residues long common to YjiK from *E. coli* K-12 and several other bacterial outer-membrane proteins, including some *Salmonella* proteins that are regulated by SdiA [137] (data not shown). This hints at a link between SdiA-mediated sensing of other bacteria and LEE-mediated type-III secretion, which, twinned with growing evidence from *Salmonella* and *Vibrio* for an influence of SdiA/LuxR on virulence and type-III secretion [136,138-141], suggests that it might be worth taking a fresh look at the influence of SdiA on LEE-mediated type-III secretion. Exhaustive PSI-BLAST searches suggest that the "SdiA-regulated domain" in fact consists of several NHL-like repeats (PF01436), which are likely to fold up into a six-bladed beta-propeller [142,143] (data not shown).

GlrR and GlrA (previously known as Orf10/L0044 and Orf11 respectively) are a pair of recently characterised regulatory proteins encoded by adjacent genes within the LEE [33,144]. Similarity has already been reported between GlrA and CaiF, a transcriptional regulator from *E. coli* and to a *Salmonella* protein [33]. PSI-BLAST searches confirm this and report similarity to several other putative transcriptional regulators (e.g. the R721-plasmid-encoded regulator YheC from *E. coli* K-12). PSI-BLAST searches with GlrR identify only one homologue outside the LEE – a protein encoded within the unfinished *Salmonella bongori* genome (data not shown). Interestingly, the *S. bongori glrR* homologue is located next to a *glrA* homologue, suggesting a functional link between the two proteins and adding weight to the finding from yeast two-hybrid studies that GlrA and GlrR interact with each other [145].

The remaining coding sequences

Six coding sequences, which initially appeared to be orphans, have recently been shown to be essential for LEE-encoded type-III secretion: Orf2, Orf4, Orf12, Orf15, Orf16 and Orf29 [33]. Thus, something can now be said about all of them based on homology and/or experimental data. PSI-BLAST searches show that Orf2 has

homologues in several other systems: *ssaE* in Spi-2, Cv2595 in Cpi-2 and *yscE* in the Ysc-Yop system (data not shown). Orf2 should thus be renamed EscE. The *ssaE* gene was found to be induced after invasion of a murine cultured cell line [146]. Like YscE in Yersinia [147], Orf2/EscE is necessary for type-III secretion through the LEE-encoded T3SS [33]. YscE interacts with TyeA [148] – thus it would seem likely that Orf2/EscE interacts with the C-terminal portion of SepL (Table 1). YscE also interacts with YscG, a protein with no detectable homologue in the LEE-encoded system [149].

Orf4, Orf12 and Orf29 possess homologues in T3SSs closely related to the LEE-encoded system, although not in the Ysc-Yop system (Table 1). None of the Orf4 or Orf29 homologues has been investigated. However, the homologue of Orf12 in the Spi-2 system, SsaM, has recently been characterised in *S. enterica* [150]. An *ssaM* deletion mutant was attenuated *in vivo* – *in vitro* it over-secreted, but failed to translocate, the effector SseJ. Furthermore, it also failed to secrete translocators. In addition, SsaM was shown to interact with another effector SpiC within the bacterial cell. The story told by this complex set of phenotypes, reminiscent of the multi-faceted YopN-like proteins or of the FliK/YscP-like proteins, stands at odds with the simpler global defect in type-III secretion reported in an *orf12* deletion mutant [33]. As with the YopN-like and FliK/YscP-like proteins, it remains to be seen whether these apparently dissimilar functional properties of homologous proteins from different T3SSs will converge into a unified mechanistic picture or represent genuine evolutionary divergence of function.

Given that Orf29 is known to interact with Orf2/EscE [145], it is tempting to speculate that, even in absence of homology, Orf29 and the chief binding partner of YscE, YscG might play similar roles. As noted above, Orf16 is a candidate for needle-length control, while the function and evolutionary origin of Orf15 remain mysterious.

Conclusion

In this survey, we have discovered several new homologous relationships and woven together threads of evidence from our own *in silico* surveys and from published experimental studies to craft numerous novel functional predictions about the LEE-encoded and other T3SSs that can now be tested in the laboratory. We invite the scientific community to test our predictions. Furthermore, in this sequence-rich, post-genomic era [2,23], we believe the time is right to roll out similar bioinformatics analyses over all known type-III secretion systems and invite others to join us in this enterprise on our new web site, 3Base, dedicated to type-III secretion <http://3base.bham.ac.uk>.

Methods

Sequences

Peptide sequences of the 44 LEE-encoded proteins from the genome-sequenced enterohaemorrhagic *E. coli* O157:H7 Sakai strain (encoded by genes ECs4548-ECs4591 in the Sakai nomenclature) were extracted from the relevant GenBank file (accession number dbj|BA000007). Analyses were performed on a Macintosh Dual 2 GHz G5 running OSX10.3.

Homology searches

PSI-BLAST searches [24] were performed on LEE-encoded proteins against a composite database that included the NCBI's non-redundant (NR) protein database and a database of predicted protein sequences from unfinished genomes (VGE-PEPT: searchable on our web site <http://www.vge.ac.uk>). To explore all possibilities, multiple PSI-BLAST searches were run for the LEE-encoded proteins using different combinations of substitution matrices (BLOSUM45, BLOSUM62, BLOSUM80), low-complexity filtering (ON, OFF), compositional-based statistics (ON, OFF), E-value inclusion threshold (E = 0.002, E = 0.02) and database (VGE-PEPT + NR, VGE-PEPT alone). PSI-BLAST results were parsed using home-produced Perl scripts and any searches that uncovered significant or near-significant matches to other type III secretion associated proteins were flagged for closer examination and potential homologues retrieved and analysed further. In some cases, PSI-BLAST searches were supplemented with iterative Hmmer searches [151] and searches of the PFAM, CDD and SMART databases [40,55,152].

Most LEE proteins produced statistically significant matches to other Type III associated proteins of known and unknown function under PSI-BLAST default conditions (except for the absence of filtering and compositional based statistics) within 1 to 5 iterations. In some cases (e.g. ECs4584/ORF5), a significant PSI-BLAST hit was dependent on an absence of low-complexity eukaryote sequences in the target database, whereas in others (e.g. ECs4572/rORF8) significant hits were only identified using the BLOSUM80 matrix.

Alignments

In most cases alignments were generated from homologues identified in PSI-BLAST searches with T-coffee under default settings [153]. In the case of the EscD alignment, the FHA-like and BON-like regions were aligned separately using T-coffee. The former was then aligned to the FHA domain profile (SM00240) from the SMART database and the latter to the BON domain profile from PFAM (PF04972), respectively, using hmalign from the Hmmer package. Similarly, the SepQ alignment was generated by aligning the COG1886 and Pfam:SpoA domain profiles with hmalign. The EscI alignment was

generated using T-coffee to align homologues parsed from the PSI-BLAST searches before the FliH sequence from *E. coli* K-12 was aligned to it using hmalign. All multiple alignments were manually curated to ensure that secondary structure elements were not broken and to minimise the number of misaligned regions (as assessed using ClustalX and T-coffee [153,154]). Alignments were coloured using default CHROMA settings [155].

Authors' contributions

MJP performed the initial PSI-BLAST and CDD searches, managed the project and wrote the paper. SAB performed exhaustive PSI-BLAST searches and other bioinformatics analyses, prepared multiple alignments and compiled these into figures. CMB prepared Figure 1.

Acknowledgements

We thank the MRC for funding SB's Bioinformatics Fellowship; we thank the Division of Infection and Immunity and the Medical School, University of Birmingham for funding CMB and the 3Base project. We thank the British Biotechnology and Biological Sciences Research Council for funding other experimental and bioinformatics work within the Pallen group on *E. coli* and type-III secretion (grant references EGA16107, D13414, 02BID08033). We thank an anonymous reviewer for pointing out the similar genetic locations of *sepD* and *yscB*.

References

- Hueck CJ: **Type III protein secretion systems in bacterial pathogens of animals and plants.** *Microbiol Mol Biol Rev* 1998, **62**:379-433.
- Pallen MJ, Chaudhuri RR, Henderson IR: **Genomic analysis of secretion systems.** *Curr Opin Microbiol* 2003, **6**:519-527.
- Cornelis GR, Van Gijsegem F: **Assembly and function of type III secretory systems.** *Annu Rev Microbiol* 2000, **54**:735-774.
- Cornelis GR: **Type III secretion: a bacterial device for close combat with cells of their eukaryotic host.** *Philos Trans R Soc Lond B Biol Sci* 2000, **355**:681-693.
- Blocker A, Komoriya K, Aizawa S: **Type III secretion systems and bacterial flagella: insights into their function from structural similarities.** *Proc Natl Acad Sci U S A* 2003, **100**:3027-3030.
- Muller S, Feldman MF, Cornelis GR: **The Type III secretion system of Gram-negative bacteria: a potential therapeutic target?** *Expert Opin Ther Targets* 2001, **5**:327-339.
- Russmann H: **Bacterial type III translocation: a unique mechanism for cytosolic display of heterologous antigens by attenuated Salmonella.** *Int J Med Microbiol* 2003, **293**:107-112.
- Russmann H: **Yersinia outer protein E, YopE. A versatile type III effector molecule for cytosolic targeting of heterologous antigens by attenuated Salmonella.** *Adv Exp Med Biol* 2003, **529**:407-413.
- Kenny B: **Mechanism of action of EPEC type III effector molecules.** *Int J Med Microbiol* 2002, **291**:469-477.
- Kaper JB, Nataro JP, Mobley HLT: **Pathogenic Escherichia coli.** *Nature Reviews Microbiology* 2004, **2**:123-140.
- McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB: **A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens.** *Proc Natl Acad Sci U S A* 1995, **92**:1664-1668.
- Jarvis KG, Giron JA, Jerse AE, McDaniel TK, Donnenberg MS, Kaper JB: **Enteropathogenic Escherichia coli contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation.** *Proc Natl Acad Sci U S A* 1995, **92**:7996-8000.
- McDaniel TK, Kaper JB: **A cloned pathogenicity island from enteropathogenic Escherichia coli confers the attaching and effacing phenotype on E. coli K-12.** *Mol Microbiol* 1997, **23**:399-407.
- Jerse AE, Yu J, Tall BD, Kaper JB: **A genetic locus of enteropathogenic Escherichia coli necessary for the production of attaching and effacing lesions on tissue culture cells.** *Proc Natl Acad Sci U S A* 1990, **87**:7839-7843.
- Hacker J, Kaper JB: **Pathogenicity islands and the evolution of microbes.** *Annu Rev Microbiol* 2000, **54**:641-679.
- Perna NT, Mayhew GF, Posfai G, Elliott S, Donnenberg MS, Kaper JB, Blattner FR: **Molecular evolution of a pathogenicity island from enterohemorrhagic Escherichia coli O157:H7.** *Infect Immun* 1998, **66**:3810-3817.
- Schauer DB, Falkow S: **Attaching and effacing locus of a Citrobacter freundii biotype that causes transmissible murine colonic hyperplasia.** *Infect Immun* 1993, **61**:2486-2492.
- Huys G, Cnockaert M, Janda JM, Swings J: **Escherichia albertii sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children.** *Int J Syst Evol Microbiol* 2003, **53**:807-810.
- Janda JM, Abbott SL, Albert MJ: **Prototypal diarrhoeagenic strains of Hafnia alvei are actually members of the genus Escherichia.** *J Clin Microbiol* 1999, **37**:2399-2401.
- Albert MJ, Faruque SM, Ansaruzzaman M, Islam MM, Haider K, Alam K, Kabir I, Robins-Browne R: **Sharing of virulence-associated properties at the phenotypic and genetic levels between enteropathogenic Escherichia coli and Hafnia alvei.** *J Med Microbiol* 1992, **37**:310-314.
- Webber C, Ponting CP: **Genes and homology.** *Curr Biol* 2004, **14**:R332-3.
- Pallen M: **From Sequence to Consequence: In silico Hypothesis Generation and Testing.** In *Methods in Microbiology Volume 33*. Edited by: Wren B and Dorrell N. London, Academic Press; 2002:27-48.
- Pallen MJ, Beatson SA, Bailey CM: **Bioinformatics, Genomics and Evolution of Non-Flagellar Type-III Secretion Systems: A Darwinian Perspective.** *FEMS Microbiol Rev* 2005, **29**:201-229.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
- Kenny B, Jepsen M: **Targeting of an enteropathogenic Escherichia coli (EPEC) effector protein to host mitochondria.** *Cell Microbiol* 2000, **2**:579-590.
- McNamara BP, Donnenberg MS: **A novel proline-rich protein, EspF, is secreted from enteropathogenic Escherichia coli via the type III export pathway.** *FEMS Microbiol Lett* 1998, **166**:71-78.
- Elliott SJ, Krejany EO, Mellies JL, Robins-Browne RM, Sasakawa C, Kaper JB: **EspG, a novel type III system-secreted protein from enteropathogenic Escherichia coli with similarities to VirA of Shigella flexneri.** *Infect Immun* 2001, **69**:4027-4033.
- Tu X, Nisan I, Yona C, Hanski E, Rosenshine I: **EspH, a new cytoskeleton-modulating effector of enterohaemorrhagic and enteropathogenic Escherichia coli.** *Mol Microbiol* 2003, **47**:595-606.
- Neves BC, Mundy R, Petrovska L, Dougan G, Knutton S, Frankel G: **CesD2 of enteropathogenic Escherichia coli is a second chaperone for the type III secretion translocator protein EspD.** *Infect Immun* 2003, **71**:2130-2141.
- Elliott SJ, O'Connell CB, Koutsouris A, Brinkley C, Donnenberg MS, Hecht G, Kaper JB: **A gene from the locus of enterocyte effacement that is required for enteropathogenic Escherichia coli to increase tight-junction permeability encodes a chaperone for EspF.** *Infect Immun* 2002, **70**:2271-2277.
- Creasey EA, Friedberg D, Shaw RK, Umanski T, Knutton S, Rosenshine I, Frankel G: **CesAB is an enteropathogenic Escherichia coli chaperone for the type-III translocator proteins EspA and EspB.** *Microbiology* 2003, **149**:3639-3647.
- Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB: **The Per regulon of enteropathogenic Escherichia coli: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler).** *Mol Microbiol* 1999, **33**:296-306.
- Deng W, Puente JL, Gruenheid S, Li Y, Vallance BA, Vazquez A, Barba J, Ibarra JA, O'Donnell P, Metalnikov P, Ashman K, Lee S, Goode D, Pawson T, Finlay BB: **Dissecting virulence: systematic and functional analyses of a pathogenicity island.** *Proc Natl Acad Sci U S A* 2004, **101**:3597-3602.

34. Park J, Teichmann SA, Hubbard T, Chothia C: **Intermediate sequences increase the detection of homology between sequences.** *J Mol Biol* 1997, **273**:349-354.
35. Park J, Karplus K, Barrett C, Hughey R, Haussler D, Hubbard T, Chothia C: **Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods.** *J Mol Biol* 1998, **284**:1201-1210.
36. Muller A, MacCallum RM, Sternberg MJ: **Benchmarking PSI-BLAST in genome annotation.** *J Mol Biol* 1999, **293**:1257-1271.
37. Jones DT, Swindells MB: **Getting the most from PSI-BLAST.** *Trends Biochem Sci* 2002, **27**:161-164.
38. Stevens FJ, Kuemmel C, Babnigg G, Collart FR: **Efficient recognition of protein fold at low sequence identity by conservative application of Psi-BLAST: application.** *J Mol Recognit* 2004.
39. Stevens FJ: **Efficient recognition of protein fold at low sequence identity by conservative application of Psi-BLAST: validation.** *J Mol Recognit* 2004.
40. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR: **The Pfam protein families database.** *Nucleic Acids Res* 2004, **32 Database issue**:D138-41.
41. Betts HJ, Chaudhuri RR, Pallen MJ: **An analysis of type-III secretion gene clusters in *Chromobacterium violaceum*.** *Trends Microbiol* 2004, **12**:476-482.
42. Marches O, Ledger TN, Boury M, Ohara M, Tu X, Goffaux F, Mainil J, Rosenshine I, Sugai M, De Rycke J, Oswald E: **Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G2/M transition.** *Mol Microbiol* 2003, **50**:1553-1567.
43. Mundy R, Petrovska L, Smollett K, Simpson N, Wilson RK, Yu J, Tu X, Rosenshine I, Clare S, Dougan G, Frankel G: **Identification of a novel *Citrobacter rodentium* type III secreted protein, EspI, and roles of this and other secreted proteins in infection.** *Infect Immun* 2004, **72**:2288-2302.
44. Garmendia J, Phillips AD, Carlier MF, Chong Y, Schuller S, Marches O, Dahan S, Oswald E, Shaw RK, Knutton S, Frankel G: **TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton.** *Cell Microbiol* 2004, **6**:1167-1183.
45. Campellone KG, Robbins D, Leong JM: **EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly.** *Dev Cell* 2004, **7**:217-228.
46. Kanack K, Crawford JA, Karmali MA, Kaper JB: **SepZ, a Novel Enteropathogenic *Escherichia coli* Type-III System Secreted and Translocated Protein: Mutation Does Not Delay Secretion of EspA/B or Translocation of Tir.** [abstract] American Society for Microbiology General Meeting, New Orleans 2004.
47. Elliott SJ, Wainwright LA, McDaniel TK, Jarvis KG, Deng YK, Lai LC, McNamara BP, Donnenberg MS, Kaper JB: **The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69.** *Mol Microbiol* 1998, **28**:1-4.
48. Kresse AU, Schulze K, Deibel C, Ebel F, Rohde M, Chakraborty T, Guzman CA: **Pas, a novel protein required for protein secretion and attaching and effacing activities of enterohaemorrhagic *Escherichia coli*.** *J Bacteriol* 1998, **180**:4370-4379.
49. Gauthier A, Puente JL, Finlay BB: **Secretin of the enteropathogenic *Escherichia coli* type III secretion system requires components of the type III apparatus for assembly and localization.** *Infect Immun* 2003, **71**:3310-3319.
50. Gauthier A, Finlay BB: **Translocated intimin receptor and its chaperone interact with ATPase of the type III secretion apparatus of enteropathogenic *Escherichia coli*.** *J Bacteriol* 2003, **185**:6747-6755.
51. Jackson MW, Plano GV: **DsbA is required for stable expression of outer membrane protein YscC and for efficient Yop secretion in *Yersinia pestis*.** *J Bacteriol* 1999, **181**:5126-5130.
52. Pallen M, Chaudhuri R, Khan A: **Bacterial FHA domains: neglected players in the phospho-threonine signalling game?** *Trends Microbiol* 2002, **10**:556-563.
53. Yeats C, Bateman A: **The BON domain: a putative membrane-binding domain.** *Trends Biochem Sci* 2003, **28**:352-355.
54. Jackson MW, Plano GV: **Interactions between type III secretion apparatus components from *Yersinia pestis* detected using the yeast two-hybrid system.** *FEMS Microbiol Lett* 2000, **186**:85-90.
55. Marchler-Bauer A, Anderson JB, Cherukuri PF, Dewese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH: **CDD: a Conserved Domain Database for protein classification.** *Nucleic Acids Res* 2005, **33 Database Issue**:D192-6.
56. Fadoulglou VE, Tampakaki AP, Glykos NM, Bastaki MN, Hadden JM, Phillips SE, Panopoulos NJ, Kokkinidis M: **Structure of HrcQB-C, a conserved component of the bacterial type III secretion systems.** *Proc Natl Acad Sci U S A* 2004, **101**:70-75.
57. Minamino T, MacNab RM: **FliH, a soluble component of the type III flagellar export apparatus of *Salmonella*, forms a complex with FliI and inhibits its ATPase activity.** *Mol Microbiol* 2000, **37**:1494-1503.
58. Koraimann G: **Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria.** *Cell Mol Life Sci* 2003, **60**:2371-2388.
59. Thunnissen AM, Dijkstra AJ, Kalk KH, Rozeboom HJ, Engel H, Keck W, Dijkstra BV: **Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography.** *Nature* 1994, **367**:750-753.
60. Thunnissen AM, Rozeboom HJ, Kalk KH, Dijkstra BW: **Structure of the 70-kDa soluble lytic transglycosylase complexed with bulgecin A. Implications for the enzymatic mechanism.** *Biochemistry* 1995, **34**:12729-12737.
61. van Asselt EJ, Kalk KH, Dijkstra BW: **Crystallographic studies of the interactions of *Escherichia coli* lytic transglycosylase Slt35 with peptidoglycan.** *Biochemistry* 2000, **39**:1924-1934.
62. Wilson RK, Shaw RK, Daniell S, Knutton S, Frankel G: **Role of EscF, a putative needle complex protein, in the type III protein translocation system of enteropathogenic *Escherichia coli*.** *Cell Microbiol* 2001, **3**:753-762.
63. Sekiya K, Ohishi M, Ogino T, Tamano K, Sasakawa C, Abe A: **Supermolecular structure of the enteropathogenic *Escherichia coli* type III secretion system and its direct interaction with the EspA-sheath-like structure.** *Proc Natl Acad Sci U S A* 2001, **98**:11638-11643.
64. Hoiczynk E, Blobel G: **Polymerization of a single protein of the pathogen *Yersinia enterocolitica* into needles punctures eukaryotic cells.** *Proc Natl Acad Sci U S A* 2001, **98**:4669-4674.
65. Sukhan A, Kubori T, Galan JE: **Synthesis and localization of the *Salmonella* SPI-1 type III secretion needle complex proteins PrgI and PrgJ.** *J Bacteriol* 2003, **185**:3480-3483.
66. Kimbrough TG, Miller SI: **Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation.** *Proc Natl Acad Sci U S A* 2000, **97**:11008-11013.
67. Kubori T, Sukhan A, Aizawa SI, Galan JE: **Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system.** *Proc Natl Acad Sci U S A* 2000, **97**:10225-10230.
68. Tamano K, Aizawa S, Katayama E, Nonaka T, Imajoh-Ohmi S, Kuwae A, Nagai S, Sasakawa C: **Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors.** *Embo J* 2000, **19**:3876-3887.
69. Blocker A, Jouihri N, Larquet E, Gounon P, Ebel F, Parsot C, Sansonetti P, Allaoui A: **Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secretion.** *Mol Microbiol* 2001, **39**:652-663.
70. Marlovits TC, Kubori T, Sukhan A, Thomas DR, Galan JE, Unger VM: **Structural insights into the assembly of the type III secretion needle complex.** *Science* 2004, **306**:1040-1042.
71. Ebel F, Podzadel T, Rohde M, Kresse AU, Kramer S, Deibel C, Guzman CA, Chakraborty T: **Initial binding of Shiga toxin-producing *Escherichia coli* to host cells and subsequent induction of actin rearrangements depend on filamentous EspA-containing surface appendages.** *Mol Microbiol* 1998, **30**:147-161.
72. Knutton S, Rosenshine I, Pallen MJ, Nisan I, Neves BC, Bain C, Wolff C, Dougan G, Frankel G: **A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells.** *Embo J* 1998, **17**:2166-2176.

73. Daniell SJ, Kocsis E, Morris E, Knutton S, Booy FP, Frankel G: **3D structure of EspA filaments from enteropathogenic Escherichia coli.** *Mol Microbiol* 2003, **49**:301-308.
74. Ide T, Laarmann S, Greune L, Schillers H, Oberleithner H, Schmidt MA: **Characterization of translocation pores inserted into plasma membranes by type III-secreted Esp proteins of enteropathogenic Escherichia coli.** *Cell Microbiol* 2001, **3**:669-679.
75. Samatey FA, Imada K, Nagashima S, Vonderviszt F, Kumasaka T, Yamamoto M, Namba K: **Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling.** *Nature* 2001, **410**:331-337.
76. Yonekura K, Maki-Yonekura S, Namba K: **Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy.** *Nature* 2003, **424**:643-650.
77. Mimori-Kiyosue Y, Vonderviszt F, Yamashita I, Fujiyoshi Y, Namba K: **Direct interaction of flagellin termini essential for polymorphic ability of flagellar filament.** *Proc Natl Acad Sci U S A* 1996, **93**:15108-15113.
78. Yoshioka K, Aizawa S, Yamaguchi S: **Flagellar filament structure and cell motility of Salmonella typhimurium mutants lacking part of the outer domain of flagellin.** *J Bacteriol* 1995, **177**:1090-1093.
79. Macnab RM: **Flagella and motility.** In *Escherichia coli and Salmonella: Cellular and Molecular Biology* Edited by: Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, et al. Washington, DC: ASM Press; 1996:123-145.
80. Homma M, DeRosier DJ, Macnab RM: **Flagellar hook and hook-associated proteins of Salmonella typhimurium and their relationship to other axial components of the flagellum.** *J Mol Biol* 1990, **213**:819-832.
81. Pallen MJ, Dougan G, Frankel G: **Coiled-coil domains in proteins secreted by type III secretion systems.** *Mol Microbiol* 1997, **25**:423-425.
82. Delahay RM, Knutton S, Shaw RK, Hartland EL, Pallen MJ, Frankel G: **The coiled-coil domain of EspA is essential for the assembly of the type III secretion translocon on the surface of enteropathogenic Escherichia coli.** *J Biol Chem* 1999, **274**:35969-35974.
83. Yip CK, Finlay BB, Strynadka NC: **Structural characterization of a type III secretion system filament protein in complex with its chaperone.** *Nat Struct Mol Biol* 2005, **12**:75-81.
84. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, Barrett SL, Cookson BT, Aderem A: **Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility.** *Nat Immunol* 2003, **4**:1247-1253.
85. Delahay RM, Frankel G: **Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited.** *Mol Microbiol* 2002, **45**:905-916.
86. Kresse AU, Beltrametti F, Muller A, Ebel F, Guzman CA: **Characterization of SepL of enterohemorrhagic Escherichia coli.** *J Bacteriol* 2000, **182**:6490-6498.
87. O'Connell CB, Creasey EA, Knutton S, Elliott S, Crowther LJ, Luo W, Albert MJ, Kaper JB, Frankel G, Donnenberg MS: **SepL, a protein required for enteropathogenic Escherichia coli type III translocation, interacts with secretion component SepD.** *Mol Microbiol* 2004, **52**:1613-1625.
88. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D: **A combined algorithm for genome-wide prediction of protein function.** *Nature* 1999, **402**:83-86.
89. Enright AJ, Ouzounis CA: **Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions.** *Genome Biol* 2001, **2**:RESEARCH0034.
90. Enright AJ, Iliopoulos I, Kyripides NC, Ouzounis CA: **Protein interaction maps for complete genomes based on gene fusion events.** *Nature* 1999, **402**:86-90.
91. Ferracci F, Day JB, Ezelle HJ, Plano GV: **Expression of a functional secreted YopN-TyeA hybrid protein in Yersinia pestis is the result of a +1 translational frameshift event.** *J Bacteriol* 2004, **186**:5160-5166.
92. Fields KA, Hackstadt T: **Evidence for the secretion of Chlamydia trachomatis CopN by a type III secretion mechanism.** *Mol Microbiol* 2000, **38**:1048-1060.
93. Sundberg L, Forsberg A: **TyeA of Yersinia pseudotuberculosis is involved in regulation of Yop expression and is required for polarized translocation of Yop effectors.** *Cell Microbiol* 2003, **5**:187-202.
94. Marenne MN, Journet L, Mota LJ, Cornelis GR: **Genetic analysis of the formation of the Ysc-Yop translocation pore in macrophages by Yersinia enterocolitica: role of LcrV, YscF and YopN.** *Microb Pathog* 2003, **35**:243-258.
95. Cheng LW, Kay O, Schneewind O: **Regulated secretion of YopN by the type III machinery of Yersinia enterocolitica.** *J Bacteriol* 2001, **183**:5293-5301.
96. Cheng LW, Schneewind O: **Yersinia enterocolitica TyeA, an intracellular regulator of the type III machinery, is required for specific targeting of YopE, YopH, YopM, and YopN into the cytosol of eukaryotic cells.** *J Bacteriol* 2000, **182**:3183-3190.
97. Iriarte M, Sory MP, Boland A, Boyd AP, Mills SD, Lambermont I, Cornelis GR: **TyeA, a protein involved in control of Yop release and in translocation of Yersinia Yop effectors.** *Embo J* 1998, **17**:1907-1918.
98. Kubori T, Galan JE: **Salmonella type III secretion-associated protein InvE controls translocation of effector proteins into host cells.** *J Bacteriol* 2002, **184**:4699-4708.
99. Matson JS, Nilles ML: **LcrG-LcrV interaction is required for control of Yops secretion in Yersinia pestis.** *J Bacteriol* 2001, **183**:5082-5091.
100. DeBord KL, Lee VT, Schneewind O: **Roles of LcrG and LcrV during type III targeting of effector Yops by Yersinia enterocolitica.** *J Bacteriol* 2001, **183**:4588-4598.
101. Journet L, Agrain C, Broz P, Cornelis GR: **The needle length of bacterial injectisomes is determined by a molecular ruler.** *Science* 2003, **302**:1757-1760.
102. Minamino T, Gonzalez-Pedrajo B, Yamaguchi K, Aizawa SI, Macnab RM: **FliK, the protein responsible for flagellar hook length control in Salmonella, is exported during hook assembly.** *Mol Microbiol* 1999, **34**:295-304.
103. Makishima S, Komoriya K, Yamaguchi S, Aizawa SI: **Length of the flagellar hook and the capacity of the type III export apparatus.** *Science* 2001, **291**:2411-2413.
104. Magdalena J, Hachani A, Chamekh M, Jouihri N, Gounon P, Blocker A, Allaoui A: **Spa32 regulates a switch in substrate specificity of the type III secretin of Shigella flexneri from needle components to Ipa proteins.** *J Bacteriol* 2002, **184**:3433-3441.
105. Geourjon C, Deleage G: **SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments.** *Comput Appl Biosci* 1995, **11**:681-684.
106. Wainwright LA, Kaper JB: **EspB and EspD require a specific chaperone for proper secretion from enteropathogenic Escherichia coli.** *Mol Microbiol* 1998, **27**:1247-1260.
107. Abe A, de Grado M, Pfuetzner RA, Sanchez-Sanmartin C, Devinney R, Puente JL, Strynadka NC, Finlay BB: **Enteropathogenic Escherichia coli translocated intimin receptor, Tir, requires a specific chaperone for stable secretion.** *Mol Microbiol* 1999, **33**:1162-1175.
108. Elliott SJ, Hutcheson SW, Dubois MS, Mellies JL, Wainwright LA, Batchelor M, Frankel G, Knutton S, Kaper JB: **Identification of CesT, a chaperone for the type III secretion of Tir in enteropathogenic Escherichia coli.** *Mol Microbiol* 1999, **33**:1176-1189.
109. Luo Y, Bertero MG, Frey EA, Pfuetzner RA, Wenk MR, Creagh L, Marcus SL, Lim D, Sicheri F, Kay C, Haynes C, Finlay BB, Strynadka NC: **Structural and biochemical characterization of the type III secretion chaperones CesT and SigE.** *Nat Struct Mol Biol* 2001, **8**:1031-1036.
110. Evdokimov AG, Tropea JE, Routzahn KM, Waugh DS: **Three-dimensional structure of the type III secretion chaperone SycE from Yersinia pestis.** *Acta Crystallogr D Biol Crystallogr* 2002, **58**:398-406.
111. Stebbins CE, Galan JE: **Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion.** *Nature* 2001, **414**:77-81.
112. van Eerde A, Hamiaux C, Perez J, Parsot C, Dijkstra BW: **Structure of Spa15, a type III secretion chaperone from Shigella flexneri with broad specificity.** *EMBO Rep* 2004, **5**:477-483.
113. Frank DW, Iglewski BH: **Cloning and sequence analysis of a trans-regulatory locus required for exoenzyme S synthesis in Pseudomonas aeruginosa.** *J Bacteriol* 1991, **173**:6460-6468.
114. Dasgupta N, Lykken GL, Wolfgang MC, Yahr TL: **A novel anti-activator mechanism regulates expression of the Pseu-**

- domonas aeruginosa type III secretion system. *Mol Microbiol* 2004, **53**:297-308.**
115. Pallen MJ, Francis MS, Futterer K: **Tetratricopeptide-like repeats in type-III-secretion chaperones and regulators.** *FEMS Microbiol Lett* 2003, **223**:53-60.
 116. Barve SS, Straley SC: **IcrR, a low-Ca²⁺(+)-response locus with dual Ca²⁺(+)-dependent functions in Yersinia pestis.** *J Bacteriol* 1990, **172**:4661-4671.
 117. Skrzyzpek E, Straley SC: **LcrG, a secreted protein involved in negative regulation of the low-calcium response in Yersinia pestis.** *J Bacteriol* 1993, **175**:3520-3528.
 118. Mecasas J, Bilis I, Falkow S: **Identification of attenuated Yersinia pseudotuberculosis strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis.** *Infect Immun* 2001, **69**:2779-2787.
 119. Bolin I, Portnoy DA, Wolf-Watz H: **Expression of the temperature-inducible outer membrane proteins of yersiniae.** *Infect Immun* 1985, **48**:234-240.
 120. Elliott SJ, Sperandio V, Giron JA, Shin S, Mellies JL, Wainwright L, Hutcheson SW, McDaniel TK, Kaper JB: **The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic Escherichia coli.** *Infect Immun* 2000, **68**:6115-6126.
 121. Haack KR, Robinson CL, Miller KJ, Fowlkes JW, Mellies JL: **Interaction of Ler at the LEE5 (tir) operon of enteropathogenic Escherichia coli.** *Infect Immun* 2003, **71**:384-392.
 122. Sperandio V, Mellies JL, Delahay RM, Frankel G, Crawford JA, Nguyen W, Kaper JB: **Activation of enteropathogenic Escherichia coli (EPEC) LEE2 and LEE3 operons by Ler.** *Mol Microbiol* 2000, **38**:781-793.
 123. Bustamante VH, Santana FJ, Calva E, Puente JL: **Transcriptional regulation of type III secretion genes in enteropathogenic Escherichia coli: Ler antagonizes H-NS-dependent repression.** *Mol Microbiol* 2001, **39**:664-678.
 124. Dorman CJ, Hinton JC, Free A: **Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria.** *Trends Microbiol* 1999, **7**:124-128.
 125. Rimsky S: **Structure of the histone-like protein H-NS and its role in regulation and genome superstructure.** *Curr Opin Microbiol* 2004, **7**:109-114.
 126. Tendeng C, Bertin PN: **H-NS in Gram-negative bacteria: a family of multifaceted proteins.** *Trends Microbiol* 2003, **11**:511-518.
 127. Beloin C, Deighan P, Doyle M, Dorman CJ: **Shigella flexneri 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein.** *Mol Genet Genomics* 2003, **270**:66-77.
 128. Deighan P, Beloin C, Dorman CJ: **Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of Shigella flexneri 2a strain 2457T.** *Mol Microbiol* 2003, **48**:1401-1416.
 129. Paytubi S, Madrid C, Fornis N, Nieto JM, Balsalobre C, Uhlin BE, Juarez A: **YdgT, the Hha paralogue in Escherichia coli, forms heteromeric complexes with H-NS and StpA.** *Mol Microbiol* 2004, **54**:251-263.
 130. Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB: **Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic Escherichia coli.** *Proc Natl Acad Sci U S A* 1999, **96**:15196-15201.
 131. Sperandio V, Torres AG, Giron JA, Kaper JB: **Quorum sensing is a global regulatory mechanism in enterohemorrhagic Escherichia coli O157:H7.** *J Bacteriol* 2001, **183**:5187-5197.
 132. Sperandio V, Li CC, Kaper JB: **Quorum-sensing Escherichia coli regulator A: a regulator of the LysR family involved in the regulation of the locus of enterocyte effacement pathogenicity island in enterohemorrhagic E. coli.** *Infect Immun* 2002, **70**:3085-3093.
 133. Sperandio V, Torres AG, Kaper JB: **Quorum sensing Escherichia coli regulators B and C (QseBC): a novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in E. coli.** *Mol Microbiol* 2002, **43**:809-821.
 134. Sperandio V, Torres AG, Jarvis B, Nataro JP, Kaper JB: **Bacteria-host communication: the language of hormones.** *Proc Natl Acad Sci U S A* 2003, **100**:8951-8956.
 135. Kanamaru K, Tatsuno I, Tobe T, Sasakawa C: **SdiA, an Escherichia coli homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohemorrhagic Escherichia coli O157:H7.** *Mol Microbiol* 2000, **38**:805-816.
 136. Ahmer BM: **Cell-to-cell signalling in Escherichia coli and Salmonella enterica.** *Mol Microbiol* 2004, **52**:933-945.
 137. Ahmer BM, van Reeuwijk J, Timmers CD, Valentine PJ, Heffron F: **Salmonella typhimurium encodes an SdiA homolog, a putative quorum sensor of the LuxR family, that regulates genes on the virulence plasmid.** *J Bacteriol* 1998, **180**:1185-1193.
 138. Michael B, Smith JN, Swift S, Heffron F, Ahmer BM: **SdiA of Salmonella enterica is a LuxR homolog that detects mixed microbial communities.** *J Bacteriol* 2001, **183**:5733-5742.
 139. Volf J, Sevcik M, Havlickova H, Sisak F, Damborsky J, Rychlik I: **Role of SdiA in Salmonella enterica serovar Typhimurium physiology and virulence.** *Arch Microbiol* 2002, **178**:94-101.
 140. Smith JN, Ahmer BM: **Detection of other microbial species by Salmonella: expression of the SdiA regulon.** *J Bacteriol* 2003, **185**:1357-1366.
 141. Henke JM, Bassler BL: **Quorum sensing regulates type III secretion in Vibrio harveyi and Vibrio parahaemolyticus.** *J Bacteriol* 2004, **186**:3794-3805.
 142. Slack FJ, Ruvkun G: **A novel repeat domain that is often associated with RING finger and B-box motifs.** *Trends Biochem Sci* 1998, **23**:474-475.
 143. Edwards TA, Wilkinson BD, Wharton RP, Aggarwal AK: **Model of the brain tumor-Pumilio translation repressor complex.** *Genes Dev* 2003, **17**:2508-2513.
 144. Lio JC, Syu WJ: **Identification of a negative regulator for the pathogenicity island of enterohemorrhagic Escherichia coli O157:H7.** *J Biomed Sci* 2004, **11**:855-863.
 145. Creasey EA, Delahay RM, Daniell SJ, Frankel G: **Yeast two-hybrid system survey of interactions between LEE-encoded proteins of enteropathogenic Escherichia coli.** *Microbiology* 2003, **149**:2093-2106.
 146. Janakiraman A, Schlauch JM: **The putative iron transport system SitABC encoded on SPII is required for full virulence of Salmonella typhimurium.** *Mol Microbiol* 2000, **35**:1146-1155.
 147. Allaoui A, Schulte R, Cornelis GR: **Mutational analysis of the Yersinia enterocolitica virC operon: characterization of yscE, F, G, I, J, K required for Yop secretion and yscH encoding YopR.** *Mol Microbiol* 1995, **18**:343-355.
 148. Swietnicki W, O'Brien S, Holman K, Cherry S, Brueggemann E, Tropea JE, Hines HB, Waugh DS, Ulrich RG: **Novel protein-protein interactions of the Yersinia pestis type III secretion system elucidated with a matrix analysis by surface plasmon resonance and mass spectrometry.** *J Biol Chem* 2004, **279**:38693-38700.
 149. Day JB, Guller I, Plano GV: **Yersinia pestis YscG protein is a Syc-like chaperone that directly binds yscE.** *Infect Immun* 2000, **68**:6466-6471.
 150. Yu XJ, Liu M, Holden DW: **SsaM and SpiC interact and regulate secretion of Salmonella Pathogenicity Island 2 type III secretion system effectors and translocators.** *Mol Microbiol* 2004, **54**:604-619.
 151. Hmmer [<http://hmmer.wustl.edu/>]
 152. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P: **SMART 4.0: towards genomic data integration.** *Nucleic Acids Res* 2004, **32 Database issue**:D142-4.
 153. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302**:205-217.
 154. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**:3497-3500.
 155. Goodstadt L, Ponting CP: **CHROMA: consensus-based colouring of multiple alignments for publication.** *Bioinformatics* 2001, **17**:845-846.
 156. Burghout P, van Boxtel R, Van Gelder P, Ringler P, Muller SA, Tommassen J, Koster M: **Structure and electrophysiological properties of the YscC secretin from the type III secretion system of Yersinia enterocolitica.** *J Bacteriol* 2004, **186**:4645-4654.
 157. Cornelis GR: **The Yersinia Ysc-Yop 'type III' weaponry.** *Nat Rev Mol Cell Biol* 2002, **3**:742-752.

158. Van Arnam JS, McMurry JL, Kihara M, Macnab RM: **Analysis of an engineered Salmonella flagellar fusion protein, FliR-FliH.** *J Bacteriol* 2004, **186**:2495-2498.
159. Lavander M, Sundberg L, Edqvist PJ, Lloyd SA, Wolf-Watz H, Forsberg A: **Proteolytic cleavage of the FliH homologue YscU of Yersinia pseudotuberculosis is essential for bacterial survival but not for type III secretion.** *J Bacteriol* 2002, **184**:4500-4509.
160. Edqvist PJ, Olsson J, Lavander M, Sundberg L, Forsberg A, Wolf-Watz H, Lloyd SA: **YscP and YscU regulate substrate specificity of the Yersinia type III secretion system.** *J Bacteriol* 2003, **185**:2259-2266.
161. Pucciarelli MG, Garcia-del Portillo F: **Protein-peptidoglycan interactions modulate the assembly of the needle complex in the Salmonella invasion-associated type III secretion system.** *Mol Microbiol* 2003, **48**:573-585.
162. Suzuki H, Yonekura K, Namba K: **Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis.** *J Mol Biol* 2004, **337**:105-113.
163. Suzuki H, Yonekura K, Murata K, Hirai T, Oosawa K, Namba K: **A structural feature in the central channel of the bacterial flagellar FliF ring complex is implicated in type III protein export.** *J Struct Biol* 1998, **124**:104-114.
164. Carrillo CD, Taboada E, Nash JH, Lanthier P, Kelly J, Lau PC, Verhulp R, Mykytczuk O, Sy J, Findlay WA, Amoako K, Gomis S, Willson P, Austin JW, Potter A, Babiuk L, Allan B, Szymanski CM: **Genome-wide expression analyses of Campylobacter jejuni NCTC11168 reveals coordinate regulation of motility and virulence by flhA.** *J Biol Chem* 2004, **279**:20327-20338.
165. Niehus E, Gressmann H, Ye F, Schlapbach R, Dehio M, Dehio C, Stack A, Meyer TF, Suerbaum S, Josenhans C: **Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori.** *Mol Microbiol* 2004, **52**:947-961.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

