

Genome-wide Screens to Identify Genes of Human Pathogenic *Yersinia* Species that are Expressed during Host Infection

Andrew J. Darwin

Department of Microbiology, New York University School of Medicine, New York, NY 10016

Abstract

An obvious goal in the study of bacteria that cause human disease is to identify the bacterial genes required for growth within the host. Historically, this has presented a significant technological challenge. However, with this goal in mind, the *in vivo* expression technology (IVET) and signature-tagged mutagenesis (STM) techniques were developed during the 1990s. These techniques have been used to identify virulence genes in the three human pathogenic *Yersinia* species, *Y. enterocolitica*, *Y. pseudotuberculosis* and *Y. pestis*, using variations of their mouse models of infection. In this review, each of these studies is described individually, including the pertinent details of how each was done, and a brief discussion of the genes identified. In addition, the results of these IVET and STM screens are compared, and the striking lack of overlap between the genes identified is discussed. Most of these studies were only recently published, which means that there have been few follow-up studies on some of the novel virulence genes identified. However, the *Y. enterocolitica hreP*, *rscR* and *psp* genes have become the subject of further studies, which are also summarized here. Finally, I briefly describe the use of the genome-wide (but not *in vivo*) technology, subtractive hybridization, to identify *Yersinia* virulence genes.

Introduction

This review is devoted to the use of “genome-wide” approaches to identify *Yersinia* virulence genes, with an emphasis on those technologies that can be directly applied during host infection. A number of these *in vivo* technologies are available. However, it is not the purpose of this review to go into the technical details of each available technique. For a comprehensive review of the available *in vivo* strategies, I refer the reader elsewhere (Handfield and Levesque, 1999).

In *Yersinia* spp, two such *in vivo* technologies have been used on multiple occasions, and these studies will be covered in detail. The first was *in vivo* expression technology (IVET), which was developed in John Mekalanos’ laboratory (Mahan *et al.*, 1993). The second was signature-tagged transposon mutagenesis (STM), developed in David Holden’s laboratory (Hensel *et al.*, 1995). In a few cases, some of the genes identified in

these screens have become the subjects of subsequent studies, which will also be summarized in this review.

I have listed most of the genes found in a separate table for each screen. In several cases, the authors of the original studies reported that analysis of some of the DNA sequence information revealed no similarity to the databases. I have decided to omit these from the tables, simply because it is not informative in the context of this review. Of course, this does not mean that I think these potentially novel genes are not interesting.

IVET screens

IVET identifies bacterial genes that are expressed during an animal infection, but not during selected laboratory growth conditions. The hypothesis is that genes that meet these criteria are likely to be required for virulence, but unlikely to be so-called housekeeping genes. However, IVET does not directly reveal whether or not the genes identified are required for virulence. For this, a null mutant must be constructed and its phenotype determined. This has been done for only a small minority of the genes identified by IVET, in a number of different bacterial species. In the case of the human pathogenic *Yersinia* spp., IVET has been used to study *Y. enterocolitica* only, in both intestinal and systemic mouse models of infection (Young and Miller, 1997; Gort and Miller, 2000). This review focuses on the human-pathogenic *Yersinia* spp. However, it is also worthy to note that IVET has also recently been used to study the fish pathogen *Y. ruckeri* (Fernandez *et al.*, 2004).

IVET identification of genes expressed early in infection

The first *Y. enterocolitica* IVET screen studied a derivative of strain 8081 (biotype 1B, serotype O8). The screen identified genes expressed in murine Peyer’s patch tissue throughout at least the first 46 hours after an orogastric infection (Young and Miller, 1997). A library of strains with random chromosomal *cat* operon fusions (encoding chloramphenicol acetyltransferase) was constructed. The source of genomic DNA for this library was a strain lacking the virulence plasmid (pYV). Whilst this was understandably done to avoid the reidentification of known pYV virulence genes, in future studies it might also be interesting to know which pYV genes would meet the subsequent selection criteria.

The *cat* operon fusion library was used to infect BALB/c mice. Chloramphenicol was administered to the animals throughout the first 46 hours of infection. This enriched for those strains with *cat* fusions expressed during infection. Strains that survived two rounds of this enrichment were recovered from the Peyer’s patches and further characterized to identify *cat* fusions that were not expressed on rich or minimal agar plates in the laboratory at 26°C. This class of strains was designated to contain *cat* fusions to host responsive elements (*hre*). 61 different

For correspondence: Department of Microbiology MSB 228, New York University School of Medicine, 550 First Avenue, New York, NY 10016
Phone: (212) 263–3223; Fax: (212) 263–8276; E-mail: darwia01@med.nyu.edu

hre allelic groups were identified, and the DNA sequence of 48 was determined in the original report (Table 1; Young and Miller, 1997).

Following the original publication, the DNA sequence of the remaining 13 was determined, and the information has been incorporated into Table 1 (G.M. Young and V. L. Miller, personal communication). In addition, some of the fusions that originally revealed no homology to the databases were analyzed further (Heusipp *et al.*, 2003). One is a fusion to a homologue of the *E. coli nadB* gene

(Table 1). Another is a fusion to a region with similarity to the *E. coli rpoE* promoter. Further sequence analysis confirmed that this is a fusion to the *Y. enterocolitica rpoE* promoter. Interestingly, attempts to construct a *Y. enterocolitica rpoE* null mutant have been unsuccessful (Heusipp *et al.*, 2003). This indicates that *rpoE* is an essential gene in *Y. enterocolitica*, as is the case in *E. coli* (De Las Penas *et al.*, 1997).

The authors of this IVET study divided the *hre* loci into functional groups (stress response, iron acquisition,

Table 1. IVET identification of *Y. enterocolitica* genes (*hre*) expressed in Peyer's patches and early in infection.

Encoded protein/homologue ^a	Predicted function/property	Role in virulence ^b
Stress response		
70% Gsh, <i>E. coli</i>	Glutathione synthesis	ND
91% YdhD, <i>E. coli</i>	Glutaredoxin	ND
56% MtpS, <i>Providencia stuartii</i>	DNA methylase	ND
69% MutL, <i>E. coli</i>	Methylation-dependent DNA repair	ND
79% HflX, <i>E. coli</i>	GTP-binding protein	ND
67% RecB, <i>E. coli</i>	Exodeoxyribonuclease V	ND
61% AcrR, <i>E. coli</i>	Stress response regulator	ND
Similarity to <i>E. coli rpoE</i> promoter	Extracytoplasmic stress response sigma factor	Essential gene
100% ClpX, <i>Y. enterocolitica</i>	ATP-binding subunit of Clp protease	ND
Iron acquisition		
100% Irp2, <i>Y. enterocolitica</i>	HMWP2, Iron acquisition	
100% Irp3, <i>Y. enterocolitica</i>	Unknown function (probable iron acquisition role)	Yes
85% FoxA, <i>Y. enterocolitica</i>	Siderophore receptor	ND
100% FyuA, <i>Y. enterocolitica</i>	Yersiniabactin receptor	Yes
85% YfuB, <i>Y. enterocolitica</i>	Iron transport	ND
59% HemD, <i>E. coli</i>	Uroporphyrinogen III synthase	ND
Cell envelope maintenance		
74% MdoG, <i>E. coli</i>	Membrane-derived oligosaccharide synthesis	ND
79% MdoH, <i>E. coli</i>	Membrane-derived oligosaccharide synthesis	Yes
66% LpxA, <i>Y. enterocolitica</i>	Acyl-transferase	Yes
Miscellaneous		
99% Tnp, <i>E. coli</i>	Transposase	ND
65% AceB, <i>E. coli</i>	Malate synthase	ND
59% CpdP, <i>Vibrio fischeri</i>	3'-5' cAMP phosphodiesterase	ND
31% HoxQ, <i>E. coli</i>	Nickel transport/Hydrogenase activity	ND
88% Tgt, <i>E. coli</i>	tRNA-guanine transglycosylase	ND
100% RscR, <i>Y. enterocolitica</i>	Transcriptional regulator	Yes
56% KpyI, <i>E. coli</i>	Pyruvate kinase	ND
100% HreP, <i>Y. enterocolitica</i>	Protease	Yes
80% NadB, <i>E. coli</i>	quinolinate synthetase, B protein	ND
92% QueA, <i>E. coli</i>	synthesis of queuine in tRNA	ND
54% MioC, <i>E. coli</i>	Unknown function	ND
81% YrbA, <i>E. coli</i>	Unknown function	ND
46% Rub, <i>Desulfovibrio vulgaris</i>	Unknown function	ND
32% Orf, <i>E. coli</i> (Accession U73857)	Unknown function	ND
<p>a. The closest homologous protein is shown, along with the percent amino acid identity. Some fusion sequences revealed no similarity to the databases, and they are omitted from this table.</p> <p>b. Indicates whether a null mutation affects <i>Y. enterocolitica</i> virulence (as measured by altered LD₅₀ or kinetics of infection). ND = not determined.</p>		

cell envelope maintenance and miscellaneous functions; Table 1). Some of the *hre* loci were known or suspected to be required for normal virulence. These include the iron acquisition proteins HMWP2 and FyuA, which are part of the high pathogenicity island (Carniel, 2001; Lesic and Carniel, 2004). However, the majority of the *hre* loci had not previously been implicated in virulence. Therefore, as part of their study, Young and Miller went on to show that null mutations in four of these novel genes (*mdoH*, *lpxA*, *hreP* and *rscR*) affected the course of a mouse infection. In addition, the *rscR* and *hreP* genes became the subjects of more extensive studies, which are described below.

***MdoH* homologue.** One of the genes studied further, originally designated as *hre-13*, is a homologue of *E. coli mdoH*. This gene is predicted to encode a protein involved in the synthesis of cyclic β -glucans (membrane derived oligosaccharides). These molecules are essential for virulence of the plant pathogen *Pseudomonas syringae* pv. *syringae* (Mukhopadhyay *et al.*, 1988), and also affect bacterial survival in environments of low osmotic strength (Kennedy, 1996). The *hre-13* null mutant was slightly delayed in the ability to colonize the Peyer's patches and mesenteric lymph nodes early in infection (Young and Miller, 1997). However, the LD₅₀ dose after oral infection was unchanged. This phenotype is consistent with a role early in the infectious process, as would be expected for genes identified in this screen.

***LpxA* homologue.** Another gene that was studied further is also predicted to be involved in cell envelope biosynthesis. The gene, originally designated *hre-14*, encodes a homologue of the acyl-transferase LpxA, which is required for lipid A synthesis (Galloway and Raetz, 1990; Vuorio *et al.*, 1991; Kelley *et al.*, 1993; Vuorio *et al.*, 1994). The *hre-14* null mutant was delayed in its ability to colonize the Peyer's patches and mesenteric lymph nodes, and this phenotype appeared to be more severe than that of the *mdoH* homologue mutant (Young and Miller, 1997). Furthermore, the *hre-14* null mutant had a slight increase in oral LD₅₀ dose (a little over 5 times that of the wild type) and the average day of death increased by approximately 25%. These phenotypes are consistent with a role throughout the course of infection.

***HreP*.** A *hreP* null mutant (originally designated *hre-22*) has a 33-fold increase in LD₅₀ following oral infection of BALB/c mice (Young and Miller, 1997). The predicted HreP protein is most similar to the cyanobacterial calcium-stimulated protease PrcA of *Anabena variabilis*. In addition, it was noted that HreP had significant similarity to several subtilisin-like proteases, which belong to the family of eukaryotic subtilisin/kexin-like proprotein convertases (Heusipp *et al.*, 2001). These proteases are initially synthesized as a single protein, consisting of the N-terminal proprotein and the C-terminal mature protease. Consistent with this, it was shown that a HreP-6 \times His fusion protein undergoes a single autocatalytic cleavage event (Heusipp *et al.*, 2001). The amino acid sequence of the cleavage site matched the consensus for the subtilisin/kexin-like proprotein convertases. However, the ability of HreP to cleave other proteins could not

be demonstrated. The authors offered some possible explanations for this, including lack of knowledge about the optimal conditions for enzyme activity, or substrate specificity, and inhibitory effects of the proprotein, which could not be purified away from the mature protease. The identification of the normal substrate for HreP holds the promise of revealing insight into its physiological function. However, this will not be trivial, especially if the substrate is a host protein. Unfortunately, conditions for *hreP* gene expression from its native promoter in the laboratory have not yet been determined. This has prevented some important studies, such as examining the subcellular location of this protease.

It is also interesting to note that *hreP* is located in a cluster of flagellar biosynthesis and chemotaxis genes (Heusipp *et al.*, 2001). These flanking genes are organized differently in enteric species. This suggests that *Y. enterocolitica* may have acquired *hreP* by horizontal transfer. The gene is also closely linked to *inv*, encoding the Invasin protein required for entry into host cells.

***RscR*.** A polar *rscR* null mutant (originally called *hre-20*) was shown to have interesting *in vivo* phenotypes. There was a modest five-fold increase in LD₅₀ following oral infection of BALB/c mice. However, when the bacterial load was monitored over time, the mutant was consistently found to be present in increased numbers in the liver and spleen when compared to wild type (Young and Miller, 1997). This kinetic phenotype was reproduced with a non-polar *rscR* in frame deletion mutant (Nelson *et al.*, 2001). Therefore, the gene designation *rscR* was chosen to indicate the effect of having the wild type gene intact (reduced splenic colonization).

RscR is predicted to be a member of the LysR family of transcriptional regulators, most closely related to the uncharacterized hypothetical YeiE protein of *E. coli* K-12 (68% identity). It was hypothesized that the *rscR* null mutant phenotype was probably due to altered expression of one or more RscR-regulated genes, which prompted a transposon-based screen to identify them (Nelson *et al.*, 2001). This led to the identification of the *rscBAC* locus, which is homologous to the *hmwABC* operon of *Haemophilus influenzae* (although the gene order is different). In the transposon screen, various *rscB::lacZ* insertion mutants were isolated, and analysis of them indicated that *rscB::lacZ* expression was induced 5 to 40-fold when RscR was overproduced.

The similarity to the *H. influenzae* operon suggested that RscA is an extracellular adhesin, which may rely on RscB and RscC for its processing and secretion. Therefore, an *rscA* null mutant was constructed and analyzed (Nelson *et al.*, 2001). The *rscA* null mutant had a similar kinetic phenotype in mice to that of the *rscR* null mutant (increased dissemination to the spleen, but normal colonization of the Peyer's patches and mesenteric lymph nodes). Although the authors conceded that their evidence was indirect, they came to the reasonable conclusion that RscR regulates the *rscBAC* locus during host infection.

It is intriguing that *rscR* or *rscA* null mutations cause increased systemic spread, but do not decrease the LD₅₀ (in fact, it may be slightly increased; Young and Miller, 1997). Further study will be required to determine the role

of the putative RscA adhesin in the pathogenesis of a *Y. enterocolitica* infection.

IVET identification of genes expressed during systemic infection

The second *Y. enterocolitica* IVET screen studied the same strain as the first (a derivative of strain 8081, which is biotype 1B, serotype O8). This screen identified genes expressed in murine spleen throughout at least the first 24 hours after an intraperitoneal infection (Gort and Miller, 2000). The same chromosomal *cat* operon fusion library from the first IVET screen was used. Once again, chloramphenicol was administered to the animals to enrich for those strains with *cat* fusions expressed during infection. Strains that survived two rounds of this enrichment were recovered from the spleen and further characterized to identify host responsive elements, exactly as described for the first IVET screen. In this case the host responsive elements identified were designated as *sif* (systemic infection factor). 31 different *sif* allelic groups were identified, and the DNA from at least one representative of each group was determined (Table 2). Subsequently, *in vivo* expression of some of the *sif* genes was confirmed by chloramphenicol-mediated enrichment of the Φ (*sif-cat*) fusion strain when it was used to infect a mouse in competition with the wild type strain (Gort and Miller, 2000).

The *sif* genes apparently encode proteins that play roles in general physiology, transcriptional regulation, and various other functions. Some of the genes were already known to play a role in virulence (*fyuA* and *manB*), or strongly suspected to do so (the *rffG* homologue). However, once again the majority of the genes identified had not previously been shown to play a role in virulence. The authors constructed a null mutation in one of these

genes, originally designated as *sif15* (see below; Gort and Miller, 2000).

There was little overlap between the genes identified in the two *Y. enterocolitica* IVET screens (compare Tables 1 and 2). In fact, the only gene identified in both screens was *fyuA*, which is part of the high pathogenicity island. The lack of overlap may indicate that different genes are expressed during early (intestinal) and late (systemic) stages of infection. It strongly suggests that there is much to be gained by doing IVET screens at different stages of infection, in different tissues, and following different modes of infection. The subject of overlap between the various IVET and STM screens will be discussed in more detail later.

Sif15. The predicted Sif15 protein is 69% identical to a putative outer membrane protein in various *Salmonella enterica* serovars, and 20% identical to HP0694 of *Helicobacter pylori*. Following a mixed intraperitoneal infection with the wild type, a *sif15* null mutant had a competitive defect in the spleen (competitive index of 0.08; see Gort and Miller, 2000). However, following orogastric infection the mutant had a significantly less severe competitive defect in the Peyer's patches (competitive index of 0.27; Gort and Miller, 2000). The authors concluded that Sif15 plays an important role during systemic infection, but is less important during colonization of the Peyer's patch tissue. A role for Sif15 during systemic infection might be expected to increase the LD₅₀, but this has not yet been tested. In the laboratory, *sif15* expression was shown to be higher at 37°C than at 26°C, consistent with a role in the host.

STM screens

STM solves the ethical, financial and labor-related concerns associated with screening large numbers of null

Table 2. IVET identification of *Y. enterocolitica* genes (*sif*) expressed in spleen during systemic infection.

Encoded protein/homologue ^a	Predicted function/property	Role in virulence ^b
37% SitC, <i>Staphylococcus epidermidis</i>	ABC transporter component	ND
80% LepA, <i>E. coli</i>	Membrane-bound GTPase	ND
76% RffG, <i>Erwinia carotovora</i>	LPS biosynthesis	ND
84% FrdA, <i>Proteus vulgaris</i>	Fumarate reductase	ND
80% MetL, <i>E. coli</i>	Aspartokinase/homoserine reductase	ND
91% YohI, <i>E. coli</i>	Putative transcriptional regulator	ND
82% BioH, <i>E. coli</i>	Biotin synthesis	ND
100% FyuA, <i>Y. enterocolitica</i>	Yersiniabactin receptor	Yes
100% ManB, <i>Y. enterocolitica</i>	O-antigen biosynthesis	Yes
78% YicD, <i>E. coli</i>	Unknown function	ND
70% YifJ, <i>Bacillus subtilis</i>	Pyruvate-flavodoxin oxidoreductase	ND
68% RhlB, <i>E. coli</i>	RNA helicase	ND
69% Orf <i>Salmonella</i>	Putative outer membrane protein	Yes
(20% HP0694 <i>Helicobacter pylori</i>)		
32% GacA <i>P. syringae</i>	Transcriptional regulator	ND

a. The closest homologous protein is shown, along with the percent amino acid identity. Some fusion sequences revealed no similarity to the databases, and they are omitted from this table.

b. Indicates whether a null mutation affects *Y. enterocolitica* virulence (as measured by altered LD₅₀ or kinetics of infection). ND = not determined.

mutants for decreased virulence in an animal model of infection (Hensel *et al.*, 1995). This is because STM allows relatively large groups of transposon-insertion mutants (e.g. 96 different mutants) to be screened in a single animal. STM will not identify essential genes because the mutants must be able to grow in the laboratory. Furthermore, mutants that can be complemented by wild type bacteria may not be identified in the mixed infections. The major advantage of STM is that it is a direct screen for decreased virulence of null mutants. It can identify mutants with either severe or subtle virulence defects, as demonstrated by the screens done with *Yersinia* spp. STM has been used in *Y. enterocolitica* (Darwin and Miller, 1999), *Y. pseudotuberculosis* (Karlyshev *et al.*, 2001; Mecsas *et al.*, 2001) and *Y. pestis* (Flashner *et al.*, 2004). Together, these studies also covered several different routes of infection (oral, intraperitoneal, intravenous and subcutaneous).

Y. enterocolitica STM screen (intraperitoneal infection)

The first *Yersinia* STM screen was done with *Y. enterocolitica*, using the same strain 8081 derivative as the IVET screens described above (Darwin and Miller, 1999). In this study, attempts to use an oral route of infection were unsuccessful. This was apparently due to the so-called "bottleneck" problem, which has been reported by a number of investigators using STM to study enteric pathogens (Mecsas, 2002). In the case of *Y. enterocolitica*, it appears that as few as 30 bacteria seed the Peyer's patches following an oral infection. Therefore, it was not possible to use a pool containing 96 different mutants in oral infections, even at a very high dose (Darwin and Miller, 1999). The effect of lowering the complexity of the pool (i.e. the number of different mutants) was not studied. However, an intraperitoneal (i.p.) route of infection, with pools of 96 mutants, was successful when a high dose was used (approximately 10^4 times greater than the LD₅₀ dose for a *Y. enterocolitica* i.p. infection).

The *Y. enterocolitica* signature-tagged transposon insertion mutants were initially isolated on minimal agar so that auxotrophs would not be present. These mutants were assembled into pools of 96 and used to infect mice by i.p. injection. Surviving bacteria were isolated from the spleen after 48 hours. Putative attenuated mutants identified from this first screen were then reassembled into new pools and screened a second time. After this double-screening procedure, attenuation was confirmed for 81% of the strains identified. This was done by detecting a defect in the ability of the mutant to compete with the wild type, when mice were infected with an equal mixture of the two strains (i.p. infection, bacteria recovered from the spleen). A total of 2015 random transposon insertion mutants were screened, and 55 attenuated mutants were identified. Subsequent DNA sequence analysis indicated that 27 different virulence loci had been identified (Table 3). Of these, nine were encoded on the virulence plasmid, and 18 on the chromosome.

Most of the virulence plasmid genes that were identified encode components of the Ysc type III secretion apparatus, or are involved in the regulation of its production. Surprisingly, the screen identified only one of the *yop* genes, which encode the effector proteins

secreted by the Ysc type III system (*yopP*, which is known as *yopJ* in *Y. pestis* and *Y. pseudotuberculosis*). The *yopP* null mutant had only a relatively subtle virulence defect (Table 3). YopP/J has been studied extensively using *in vitro* models, where it has been shown to play a role in inducing host cell apoptosis (reviewed by Orth, 2002). The reason(s) why more *yop* genes were not identified in the screen is unknown. One intriguing possibility is that some *yop* mutations can be complemented by a corresponding *yop*⁺ strain during the mixed infections. However, this has not been tested experimentally. A less interesting, but perhaps more likely possibility is that the transposon did not insert randomly in the virulence plasmid. The failure to identify *yop* mutants is a common theme for all of the *Yersinia* STM studies (see below).

Nine of the 18 chromosomal virulence loci are in a single locus that is involved in biosynthesis of the O-antigen component of lipopolysaccharide (Table 3). All of these mutants had significant virulence defects as measured by the competition assay following intraperitoneal infection. This is consistent with the observation that a spontaneous *Y. enterocolitica* O-antigen mutant has a significantly increased LD₅₀ dose when administered by the oral route of infection (Zhang *et al.*, 1997). Furthermore, the O-antigen is known to be important for a number of enteric pathogens, and O-antigen mutants had previously been isolated in STM screens of *S. typhimurium* and *Vibrio cholerae* (Hensel *et al.*, 1995; Chiang and Mekalanos, 1998). *Y. pseudotuberculosis* O-antigen mutants were also isolated in a subsequent STM screen (see below).

The remaining nine chromosomal loci are predicted to be involved in a variety of functions (Table 3). These include the biosynthesis of cell envelope components (*yifH* and *nlpD*), phosphate (*pstC*) and iron (*irp1*) acquisition, and stress response (*dnaJ*). Two independent mutants had transposon insertions in a homologue of the *E. coli* *yibP* gene. This gene has no known function, but may be co-transcribed with an upstream gene encoding a putative 2,3-diphosphoglycerate-independent phosphoglyceromutase. These two genes are conserved in the same order in *Pseudomonas syringae* pv. *tomato*, in which a transposon insertion in the upstream gene causes attenuation in a tomato plant infection (Morris *et al.*, 1995). It is interesting that mutations in homologous loci, which might be involved in carbon metabolism, cause attenuation in both animal and plant models of infection, without affecting growth *in vitro*. It is also interesting to note that the *E. coli* YibP protein is approximately 40% identical to the C-terminal domain of NlpD, also identified in the STM screen.

When the *Y. enterocolitica* STM screen was originally published, one of the chromosomal loci was found to have no significant homology to database entries (Darwin and Miller, 1999). However, with the completion of the *Y. enterocolitica* genome sequence, I have re-examined the site of this transposon insertion. It is now clear that the transposon is inserted into the *Y. enterocolitica* *yspC* gene, which encodes one of the secreted effectors of the recently discovered Ysa type III secretion system (Foultier *et al.*, 2003). The *yspC* null mutant had the most subtle virulence defect of all of the mutants isolated in the screen. However, this is consistent with the subtle

Table 3. <i>Y. enterocolitica</i> virulence genes identified by STM after intraperitoneal infection and recovery of the bacteria from the spleen.		
Encoded protein/homologue	Function or property	Competitive index ^b
<i>Virulence plasmid</i>		
<i>yscU</i> (100%, Ye)	Ysc-Yop Type III secretion	0.00025
<i>lcrV</i> (100%, Ye)	Ysc-Yop Type III secretion	0.000079
<i>yscR</i> (100%, Ye)	Ysc-Yop Type III secretion	0.0043
<i>yscC</i> (100%, Ye)	Ysc-Yop Type III secretion	0.00071
<i>yscL</i> (100%, Ye)	Ysc-Yop Type III secretion	0.00053
<i>virF</i> (100%, Ye)	Yop regulon transcriptional activator	0.00011
<i>virG</i> (100%, Ye)	Ysc-Yop Type III secretion	ND
<i>yopP</i> (100%, Ye)	Yop effector protein	0.17
<i>sycT-yopM</i> (100%, Ye)	Intergenic region	0.000031
<i>O-antigen</i>		
<i>dhA</i> (100%, Ye)	Glucose-1-phosphate cytidyltransferase	0.00013
<i>dhB</i> (100%, Ye)	CDP-glucose 4,6-dehydratase	< 0.00011
<i>wbcC</i> (100%, Ye)	Abequosyltransferase	0.0065
<i>wbcF</i> (100%, Ye)	Unknown	< 0.081
<i>wbcH</i> (100%, Ye)	Galactoside 2-L-fucosyltransferase	< 0.000056
<i>wbcI</i> (100%, Ye)	Galactosyltransferase	0.000078
<i>manC</i> (100%, Ye)	GDP-mannose pyrophosphorylase	< 0.00038
<i>manB</i> (100%, Ye)	Phosphomannomutase	0.00011
<i>galE</i> (100%, Ye)	UDP-glucose 4-epimerase	< 0.011
<i>Miscellaneous chromosomal mutants</i>		
<i>yifH</i> (67%, Ec)	Enterobacterial common antigen synthesis	0.07
<i>dnaJ</i> (79%, St)	Heat shock response	0.0052
<i>pstC</i> (72%, Ec)	Inorganic phosphate importer	0.005
<i>topA</i> (93%, Ec)	DNA topoisomerase I	0.0068
<i>nlpD</i> (72%, Ec)	Outer membrane lipoprotein	0.10
<i>pspC</i> (62%, Ec)	Regulation of phage shock protein operon	0.000041
<i>irp1</i> (100%, Ye)	Siderophore synthesis	0.017
<i>yibP</i> (40–55%, Ec)	Unknown	0.18
<i>yspC</i> (100% Ye)	Ysa type III secretion system effector	0.36
<p>a. Amino acid identity (over the region sequenced) compared to the most similar homologue (Ye, <i>Yersinia enterocolitica</i>; Ec, <i>Escherichia coli</i>; St, <i>Salmonella typhimurium</i>).</p> <p>b. Mice were infected i.p. with an input ratio of approximately 1:1 (mutant: wild type bacteria). Survivors were recovered from the spleen after 48 hours, and the output ratio of mutant to wild type was determined. Competitive index is the output ratio divided by input ratio. A competitive index of less than one indicates that the mutant is less virulent than the wild type. The lower the competitive index, the more severe is the virulence defect. A number beginning with "<" indicates that no mutant bacteria were recovered in one or more of the test animals. ND = not determined. In some cases multiple mutations in the same gene were isolated. In this case, the mutant with the lowest competitive index is shown.</p>		

effect of a *ysaV* null mutation on virulence following either oral or i.p. infections (Haller *et al.*, 2000). Together, all of these observations support a role for the Ysa type III secretion system in the pathogenesis of a *Y. enterocolitica* infection.

One of the chromosomal virulence genes identified is homologous to the *pspC* gene of *E. coli*. The *Y. enterocolitica pspC* mutation caused a level of attenuation that was equivalent to that of a virulence plasmid-cured strain. Essentially, the *pspC* mutant is completely avirulent as measured by the i.p. infection competition assay. The *Y. enterocolitica psp* locus became the subject of a subsequent study (see below), and continues to be an area of interest in my laboratory.

The Psp system. The *E. coli* K-12 phage shock protein (*psp*) locus was discovered because infection with a filamentous phage caused massive production of the PspA protein (Brissette *et al.*, 1990). Subsequent work demonstrated that PspA protein synthesis is induced by the mislocalization of a secretin protein, which is an outer membrane pore-forming protein used by the phage to secrete progeny from the infected cell. Homologous secretins are also essential components of a number of bacterial systems, including type II and type III secretion systems, and type IV pilus biosynthesis. The overproduction of a number of these secretin proteins has been shown to induce *E. coli* PspA synthesis. The role of the *E. coli* Psp system is unknown, but it has been hypothesized to be a stress response system that is

activated by an unknown signal (reviewed by Model *et al.*, 1997; Darwin, 2005).

The *Y. enterocolitica* *pspC* null mutant was avirulent in mice, but grew normally in LB broth at 26°C (Darwin and Miller, 1999). However, subsequent analysis indicated that the mutant had a growth defect under conditions that induce production of the virulence-plasmid encoded Ysc type III secretion system. Specifically, mislocalization of the YscC secretin protein caused a complete growth arrest of the *pspC* mutant (Darwin and Miller, 2001). Therefore, it was concluded that the attenuation of the *pspC* mutant was due, at least in part, to stress resulting from mislocalization of the YscC secretin.

This conclusion assumes that a certain proportion of YscC protein does not become correctly localized in the outer membrane during host infection. This has not been directly tested, and it would not be easy to do so. However, it is somewhat reminiscent of the situation involving the assembly of the P pilus of uropathogenic *E. coli*. In this case, it is proposed that some P pilus subunits become mislocalized. This is sensed by the Cpx extracytoplasmic stress response pathway, which itself can control P pilus biosynthesis (Hung *et al.*, 2001). However, whilst the Psp extracytoplasmic stress response system somehow senses mislocalized YscC protein, there is no evidence that the *psp* system directly controls biosynthesis of the Ysc type III secretion system itself. In fact, the *Y. enterocolitica* *pspC* null mutant still assembles a functional Ysc system, at least *in vitro* (Darwin and Miller, 2001).

Analysis of the *Y. enterocolitica* Psp system is ongoing in my laboratory, both with respect to its role in virulence, and its role in the physiology of the bacterial cell. We are attempting to understand the nature of the inducing signal (Maxson and Darwin, 2004), whether other virulence factors besides YscC might also be inducers, the molecular details of the signal transduction mechanisms, and to identify genes controlled by the Psp response system (Green and Darwin, 2004).

Y. pseudotuberculosis STM screen (oral infection)

The results from two *Y. pseudotuberculosis* STM screens were reported in 2001, both of which studied the same strain (strain YPIII pIB1; Karlyshev *et al.*, 2001; Meccas *et al.*, 2001). The first of these was able to overcome the “bottleneck” problem reported above, which allowed the oral route of infection to be used (this was achieved by reducing the complexity of the mutant pool to 48 different mutants). In an informative set of preliminary experiments, it was concluded that oral infection with a pool of 48 mutants would allow the identification of those mutants unable to survive in the cecum, and possibly the mesenteric lymph nodes. However, subsequent dissemination to the spleen was too inefficient to allow the identification of attenuated mutants in this tissue (Meccas *et al.*, 2001).

Another series of preliminary experiments indicated that a strain unable to secrete the Yops was not complemented by the wild type strain in mixed infections. However, this does not rule out the possibility that a defect in only one specific Yop protein might be complemented by the wild type strain. Of course, addressing this question for each of the secreted Yops would be a significant

undertaking, beyond the scope of their study. However, it's an interesting question for the future.

The authors of this study also reported that, following oral infection, the *Y. pseudotuberculosis* Yop-secretion mutant grew less well than the wild type in all tissues, and was most significantly attenuated in the cecum, Peyer's patches and spleen. However, the Yop secretion mutant was occasionally detected in the mesenteric lymph nodes, where the fold enrichment of the wild type over the mutant was occasionally less than 20-fold (Meccas *et al.*, 2001).

Their preliminary experiments led the authors to conclude that the cecum may be the best selective environment to identify attenuated mutants following orogastric infection. Therefore, a library of 960 *Y. pseudotuberculosis* signature-tagged transposon insertion mutants was generated and screened for reduced virulence following oral infection of BALB/c mice. After a double-screening procedure, similar to that described above for *Y. enterocolitica*, 19 mutants were found that were not detected in the cecum and/or the Peyer's patches and mesenteric lymph nodes. However, only 13 of these 19 mutants had a virulence defect when they were tested individually in separate oral mouse infections. In particular, it was found that mutants that were absent from the Peyer's patches and mesenteric lymph nodes, but present in the cecum in the original screening procedure, were as virulent as wild type when tested individually. Therefore, the authors concluded that their screen was only able to reliably identify mutants defective for survival in the cecum.

DNA sequence analysis of the transposon-insertion sites in the attenuated mutants revealed that 13 different virulence loci had been identified (Table 4). Of these, six were on the virulence plasmid, and seven on the chromosome. Four of the virulence plasmid mutations affected structural components of the Ysc type III secretion system (*yscH*, *yscU*, *yscB* and *yscL*), and these mutants were unable to secrete any of the Yops (Meccas *et al.*, 2001). One of the virulence plasmid insertions was in the *IcrV* gene, and this mutant secreted all of the Yops except for LcrV, YopB and YopD. An *IcrR* mutant was also isolated, but it behaved like the wild type strain *in vitro*, secreting all Yops under low-calcium conditions, but not under high-calcium conditions. The isolation of these virulence plasmid mutants, and the preliminary experiments with the Yop-secretion mutant, indicate that the Ysc type III secretion system is important for survival in the cecum. More specifically, it suggests that the secretion of one or more secreted Yop proteins plays a role in the cecum, which is an intriguing observation. However, no individual *yop* insertion mutants were isolated in the screen. This is similar to the *Y. enterocolitica* STM screen, in which only one *yop* gene mutant was identified amongst many mutations affecting structural components of the Ysc system, or its regulation (Darwin and Miller, 1999).

Three of the chromosomal mutations affected biosynthesis of the O-antigen component of lipopolysaccharide (Table 4). Surprisingly, these mutants were defective in the ability to invade epithelial cells *in vitro* (Meccas *et al.*, 2001). In fact, the invasion defect of the O-antigen mutants was similar to that of an invasin-

Table 4. <i>Y. pseudotuberculosis</i> virulence genes identified by STM after oral infection and recovery of the bacteria from the cecum.		
Encoded protein/homologue ^a	Function or property	Cecum virulence defect ^b
<i>Virulence plasmid</i>		
<i>yscH</i>	Ysc-Yop Type III secretion	ND
<i>yscU</i>	Ysc-Yop Type III secretion	ND
<i>yscB</i>	Ysc-Yop Type III secretion	ND
<i>yscL</i>	Ysc-Yop Type III secretion	ND
<i>lcrR</i>	Ysc-Yop Type III secretion	- 2.01
<i>lcrV</i>	Ysc-Yop Type III secretion	ND
<i>O-antigen</i>		
<i>ddhC</i>	Glucose-1-phosphate cytidyltransferase	ND
<i>wzx</i>	O-antigen flippase	- 2.71
<i>gmd</i>	GDP-D-mannose dehydratase	- 2.97
<i>Miscellaneous chromosomal mutants</i>		
<i>inv</i>	Invasin, entry into host cells	- 1.63
<i>suffl</i>	Unknown function (peptidoglycan biosynthesis?)	ND
<i>cls</i>	Cardiolipin synthesis	- 3.42
<i>ksgA</i>	Ribosomal protein, kasugamycin resistance	- 0.59
<p>a. For homologous genes, the level of amino acid identity, or the organism with the homologue, was not reported.</p> <p>b. Mice were infected with an equal mixture of mutant and wild type strains. 5 days later, the numbers of wild type and mutant bacteria in the cecum was determined. The virulence defect is expressed as: log CFU of mutant – log CFU of wild type.</p>		

defective mutant (*inv*), which was also isolated in the screen. However, it is clear that this invasion defect is not the only deficiency of the O-antigen mutants, because they were severely attenuated in an i.p. infection, whereas the *inv* mutant was fully virulent (Mecses *et al.*, 2001). Furthermore, O-antigen mutants were also isolated in the *Y. enterocolitica* STM screen, which used an i.p. route of infection (Darwin and Miller, 1999).

The remaining four chromosomal mutations affected *inv* (described above), and three other genes that had not previously been implicated in virulence (*ksgA*, *suffl* and *cls*). None of these latter three mutants were defective in Yop synthesis or delivery, or in the ability to invade cultured epithelial cells *in vitro*. KsgA is a ribosomal protein that is targeted by the drug kasugamycin. The *ksgA* mutant had a growth defect under some conditions, which might explain its virulence defect (although, mice shed the *ksgA* mutant 40 days postinfection, suggesting that it survives and replicates in the cecum for long periods). The authors also discussed the possibility that the virulence phenotype of the *ksgA* mutant could be due to polar effects on downstream genes, which are predicted to be involved in metal ion transport.

The *suffl* and *cls* mutations are both likely to affect the cell envelope. The *E. coli* *suffl* gene was originally identified as a multicopy suppressor of a temperature-sensitive *ftsI* (PBP-3) mutation. Although the role of *suffl* has not been studied, it seems likely that it plays a role in peptidoglycan biosynthesis. The *cls* gene is predicted to encode cardiolipin synthase, which makes one of the primary phospholipids in the cell envelope.

Finally, the authors of this study made two additional observations worthy of note. First, 12 of the 13 mutants identified as attenuated in the cecum were also defective for growth in the spleen following an i.p. infection. This

indicates that these mutants would have been isolated regardless of the route of infection used in the screen. However, there was one exception (the *inv* mutant), which suggests some benefit to doing STM screens using different routes of infection with the same pathogen-animal system. Second, in their preliminary experiments the authors reported that they occasionally found signature-tagged strains that were present in the spleen, but absent from the mesenteric lymph nodes of the same mouse. It is possible that these strains passed through the mesenteric lymph nodes but did not establish a persistent infection there, or that they were able to infect the spleen without passing through the mesenteric lymph nodes at all. These observations point to the usefulness of signature-tagged strains in studying the dynamics of an animal infection, in addition to their use in identifying attenuated mutants.

Y. pseudotuberculosis STM screen (intravenous infection)

The authors of the second *Y. pseudotuberculosis* STM screen made some modifications to the technique (Karlyshev *et al.*, 2001). First, each transposon had two different signature tags, which had been pre-selected on the basis of uniform hybridization efficiency, and lack of cross hybridization. Second, a high-density oligonucleotide array was used for the detection of signature tags, which allowed quantification of the hybridization signals. An intravenous route of infection was used (tail vein injection). A dose 30 to 300 times higher than the LD₅₀ was used to avoid any potential bottleneck problem.

A library of 603 signature-tagged transposon insertion mutants was generated and used to infect mice in pools of 30 or 60 different mutants. Surviving bacteria were recovered from the spleen after three days. From this, 31 putative attenuated mutants were identified, which were in

30 different loci. This is a higher percentage of attenuated mutants than occurred in the other *Yersinia* STM screens, as well as some of the STM screens in other species. The authors suggested that this might have been the result of their modifications to the STM technique, which allowed quantification of hybridization signals and increased overall reliability due to the double tags. However, it should also be noted that a lower stringency of criteria was used to define attenuation in their study when compared to the other *Yersinia* STM screens. Of the 31 different mutants, the competitive index of 20 was determined individually. Of these, the authors noted that only 14 had a competitive index of less than 0.3. Of these 14 mutants, two had an *in vitro* competitive defect that was the same or even more severe than the competitive defect in mice. Therefore, if the criteria used to define attenuation were the same

as in the other *Yersinia* studies (see above), then the percentage of attenuated mutants isolated in this screen would probably be quite similar.

Twenty-seven of the putative virulence loci that were identified in the screen are listed in Table 5. However, as mentioned above, it is not clear if all of these meet the criteria used in the other *Yersinia* STM screens. The virulence defect of each individual mutant would have to be determined in order to clarify this.

Many of the genes identified are predicted to be involved in lipopolysaccharide biosynthesis, which is consistent with the other *Yersinia* STM screens. However, unlike the other STM screens, none of the mutations were in the virulence plasmid. As the authors pointed out, this is probably because the transposon mutant library was pre-screened on Congo red magnesium oxalate plates to

Encoded protein/homologue ^a	Function or property	Competitive index ^b
<i>LPS biosynthesis</i>		
YPO0054, 97%	Glycosyltransferase	0.03*
YPO1382, 97%	LpsA, glycosyltransferase	0.08*
YPO2174, 98%	UDP-glucose-6-dehydrogenase	ND
YPO3099, 96%	ManC, mannose-1-P guanylyltransferase	0.43*
YPO3100, 98%	Fcl, fucose synthetase	0.13
YPO3104, 90%	O-antigen polymerase	0.29
YPO3114, 98%	DdhB, CDP-D-glucose-dehydratase	ND
YPO3116, 95%	AscD, ascarylose biosynthesis	0.04
<i>S. enterica</i> Wzx, 80%	O-antigen flippase	0.003
<i>Miscellaneous chromosomal mutants</i>		
YPO0702, 99%	Putative lipoprotein	ND
YPO1108, 98%	Citrate synthase	0.48*
YPO1174, 96%	Putative adhesin	0.53
YPO1186, 98%	Amino acid transport	0.055
YPO1987, 95%	Unknown function	0.021
YPO1994, 97%	Unknown function	0.084
YPO2287, 96%	Amino acid transport	0.0036
YPO2440, 98%	Iron transport	0.25
YPO2532, 100%	Unknown function	ND
YPO2712, 98%	RseA, negative regulation of <i>rpoE</i>	ND
YPO3004, 97%	Prodipeptidase	0.27
YPO3144, 97%	MdIB, multidrug resistance protein	0.21
YPO3572, 98%	Putative transcriptional regulator	0.44
YPO3657–8, 97%	Unknown function	ND
YPO3834, 99%	PldA, phospholipase A	0.017
YPO3965, 96%	VirA, His kinase	0.41
<i>Xyella fastidiosa</i> orf, 70%	Phage-related transcriptional activator	ND
Phage HP1 orf, 59%	Unknown	0.89
<p>a. Amino acid identity (over the region sequenced) compared to the most similar homologue. In most cases, this was a <i>Y. pestis</i> CO92 orthologue (YPO). Where two genes are listed, the transposon insertion was in the intergenic region. Some transposon insertions were in regions with no sequence similarity to the databases, and they are omitted from this table.</p> <p>b. Mice were infected i.v. with an input ratio of approximately 1:1 (mutant: wild type bacteria). Survivors were recovered from the spleen after three days, and the output ratio of mutant to wild type was determined. Competitive index is the output ratio divided by input ratio. A competitive index of less than one indicates that the mutant is less virulent than the wild type. The lower the competitive index, the more severe is the virulence defect. ND = not determined. An asterisk indicates mutants where the <i>in vitro</i> competitive index (not shown) was equal or less than this mouse competitive index. In cases where more than one mutation in the same gene was isolated, the mutant with the lowest competitive index is shown.</p>		

confirm the presence of the virulence plasmid. Disruption of structural or regulatory components of the Ysc type III secretion system would probably cause an abnormal phenotype on these plates, resulting in their elimination from the mutant library. However, insertions in *yop* genes would be unlikely to cause an abnormal phenotype on magnesium oxalate plates (for example, a *yopP* null mutant has a wild type phenotype; see Darwin and Miller, 1999). Therefore, the pre-screening does not explain the fact that individual *yop* mutants were not identified in this STM screen. However, this result is consistent with the rarity of individual *yop* mutants in the other *Yersinia* STM screens.

The other genes identified in the screen are predicted to encode a variety of different functions. These include amino acid and iron transport proteins, transcriptional regulators and a putative adhesin. One of the transposon insertions disrupted the *rseA* gene, which encodes a negative regulator of the RpoE sigma factor. The *Y. enterocolitica* *rpoE* promoter was identified in an IVET screen (Table 1). Together, these observations suggest that modulation of the *Yersinia* RpoE extracytoplasmic stress response is important during host infection.

A phospholipase gene was also identified in this screen. Phospholipases had been identified as virulence determinants in other bacterial pathogens, but not in *Y. pseudotuberculosis*. Therefore, the authors went on to characterize the phospholipase mutant in more detail, as discussed below.

During the characterization of the transposon insertion sites, the authors noted whether or not there was an orthologue of each gene in *Y. pestis* (Table 5; Karlyshev *et al.*, 2001). Most of the genes identified did have a *Y. pestis* orthologue. However, there were some exceptions. The authors suggested that these exceptions might contribute to the different tropisms of the closely related *Y. pseudotuberculosis* and *Y. pestis* pathogens, which are predicted to have had a relatively recent common ancestor (Achtman *et al.*, 1999).

PldA. One of the attenuated mutants had a transposon insertion in a gene encoding a 282 amino acid protein with homology to a family of bacterial outer membrane phospholipases A. The mutant had a mouse competitive index of 0.017, indicating significant attenuation, and an *in vitro* competitive index of 0.46. Furthermore, following i.v. injection the median lethal dose of the mutant was approximately 200-fold higher than that of the wild type strain (Karlyshev *et al.*, 2001). Note that the median lethal dose is the expected median dose required to produce morbidity or death. This is different to the LD₅₀, which measures the dose expected to kill 50% of infected animals.

The *pldA* mutant was shown to have significantly reduced phospholipase activity (172 units) when compared to the wild type (449 units). The remaining activity was postulated to be due to other phospholipases. Indeed, a non-homologous phospholipase A (YpIA) has been characterized in *Y. enterocolitica*, and implicated in its virulence (Schmiel *et al.*, 1998). PCR analysis indicated that a homologous *ypIA* gene is also present in *Y. pseudotuberculosis* (Karlyshev *et al.*, 2001), possibly

explaining the residual activity of the *pldA* mutant. In a previous study the *ypIA* gene was not detected in *Y. pseudotuberculosis* by southern hybridization analysis with a *Y. enterocolitica* *ypIA* probe (Schmiel *et al.*, 1998). However, this may simply have been due to the level of stringency used in the experiment.

Phospholipases C are known to be important virulence factors for a number of bacterial pathogens (Titball, 1993). However, the role of phospholipases A in virulence is less well studied, with YpIA of *Y. enterocolitica* being an exception. The authors postulated that PldA might mimic the effects of mammalian phospholipases A, which release arachidonic acid from host cell membrane phospholipids. The released arachidonic acid can serve as a substrate for the generation of a variety of inflammatory mediators.

It was also postulated that PldA might play a role in the invasion of host cells, although this was not tested. The authors based this hypothesis on two observations. First, PldA of *Helicobacter pylori* is essential for the colonization of gastric mucosa (Dorrell *et al.*, 1999). Second, activation of host phospholipase A2 in cultured epithelial cells was required for invasion by *S. enterica* serovar *Typhimurium* (Pace *et al.*, 1993). Although this involves a host phospholipase A, it is possible that the surface-bound PldA of *Y. pseudotuberculosis* could contribute to this process. This is an interesting hypothesis that will have to be tested in future experiments.

Y. pestis STM screen (subcutaneous infection)

A relatively small scale STM screen was reported for *Y. pestis* (Flashner *et al.*, 2004). In this case a subcutaneous (s.c.) route of infection was used to mimic a natural infection from a flea bite. The *Y. pestis* Kimberley53 strain was used, which was obtained by passage of the Kimberley strain (Ben-Gurion and Hertman, 1958) in mice. The LD₅₀ of the Kimberley53 strain following subcutaneous infection of mice was as low as 1 CFU.

Three hundred *Y. pestis* Kimberley53 signature-tagged transposon insertion mutants, in pools of 20 mutants each, were screened by s.c. infection of female OF1 outbred mice with a total dose of 10⁴ CFU. Mice were sacrificed 48 hours post-infection and surviving bacteria were recovered from the spleen. After a two-round screening procedure, similar to those described above, 16 putative attenuated mutants were identified. All of these mutants competed equally with the wild type strain when grown in rich media *in vitro*. The mutants were divided into four different groups, depending on their virulence properties (Table 6). Some of the mutants had indistinguishable virulence properties from the wild type. However, the authors reported that they were consistently unable to compete in the spleen when administered in a pool of at least 10 different mutants. Therefore, they are included in Table 6.

As with the other *Yersinia* STM screens, the disrupted genes encode a variety of functions. An apparently striking difference from the other screens is the lack of O-antigen biosynthesis mutants. However, the authors pointed out that a trivial explanation for this is that the O-antigen biosynthesis genes are inactive in *Y. pestis*. Only one

Table 6. <i>Y. pestis</i> virulence genes identified by STM after subcutaneous infection and recovery of the bacteria from the spleen.		
Encoded protein ^a	Function or property	Competitive index ^b
<i>Mouse non-lethal</i> ^c		
YPO3728	PurH, purine biosynthesis	< 10 ⁻⁷
YPCD1.49	LcrF, regulator of Yop regulon	< 10 ⁻⁷
YPO3357	Pcm, protein-L-isoaspartate O-methyl transferase	3 × 10 ⁻⁵
<i>Delayed mouse lethality</i> ^c		
YPO3077	PurK, purine biosynthesis	0.002
YPO3505	GreA, Transcription elongation factor	0.08
<i>Unchanged mouse lethality (CI < 0.3)</i> ^c		
YPMT1.89	DnaE, DNA polymerase III α subunit	0.035
YPO2027	Putative sulfate transporter	0.054
YPO0458–9	750 p upstream of <i>thrA</i> (asparatokinase)	0.057
YPO3045–6	Between convergently transcribed genes	0.15
YPO1179	Putative nuclease	0.26
<i>Unchanged mouse lethality (CI > 0.3)</i> ^c		
YPO3973	Putative metalloprotease	0.52
YPO1994	Unknown function	0.56
YPO2793	Putative membrane protein, unknown function	0.78
YPO2947	Unknown function	1.07
YPO2471	Putative exported protein, unknown function	1.54
YPO1003–4	Upstream of putative autotransporter	1.87
<p>a. This <i>Y. pestis</i> CO92 orthologue (YPO) of the <i>Y. pestis</i> Kimberley53 gene is shown. Where two genes are listed, the transposon insertion was in the intergenic region.</p> <p>b. Mice were infected subcutaneously with an input ratio of approximately 1:1 (mutant: wild type bacteria). Survivors were recovered from the spleen after 48 hours, and the output ratio of mutant to wild type was determined. Competitive index (CI) is the output ratio divided by input ratio. A competitive index of less than one indicates that the mutant is less virulent than the wild type. The lower the competitive index, the more severe is the virulence defect. ND = not determined.</p> <p>c. Groups of three mice were infected. Mouse non-lethal mutants did not cause death in any of three infected mice subcutaneously with 100 CFU of each mutant, and monitored for 21 days. All three mice survived for mouse non-lethal mutants, and all three mice died for all other classes. The mean time to death was longer than for the wild type strain in the case of the delayed mouse lethality strains.</p>		

virulence plasmid mutant was obtained (*IcrF*), but this is probably a reflection of the relatively small mutant library.

With a goal of identifying a possible vaccine strain, the authors focused their attention on a mutant with an insertion in *pcm*, encoding a protein-L-isoaspartate O-methyl transferase. In *E. coli*, Pcm is important for survival in a variety of stress conditions (Li and Clarke, 1992; Visick *et al.*, 1998). This *Y. pestis pcm* mutant was unable to kill mice and had a severe competitive defect in the spleen (Flashner *et al.*, 2004). The authors reported that it was superior to the EV76 live vaccine strain, inducing 10- to 100-fold higher antibody titers to the protective V and F1 antigens and because it was better at conferring protective immunity.

IVET and STM: comparison and conclusions

The IVET and STM techniques have now been used to study a number of different pathogens. Therefore, it seems appropriate to consider the question of how the two compare. For this review, I will confine this comparison to the case of *Y. enterocolitica* where both have been used. The most striking observation is the almost complete lack of overlap between the genes identified (compare Tables 1 – 2 with Table 3). This is true,

even when only comparing the IVET and STM studies in which the i.p. route of infection was used, and bacteria were harvested from the spleen after two days (Tables 2 and 3). However, there is one important caveat to this observation. Both of the *Y. enterocolitica* IVET screens used operon fusion libraries generated from the genomic DNA of a virulence plasmid-cured strain. This was done so that novel virulence genes might be identified, rather than the known virulence plasmid genes. If the virulence plasmid DNA had been included in the library, it seems likely that some of the genes would have been identified in the IVET screens, which would increase the amount of overlap with the STM screen.

Aside from the special case of virulence plasmid genes, why was there so little overlap in the genes identified by the two techniques? One reason is that the techniques ask different questions. The IVET screens were designed to identify genes expressed in the animal, but not in the laboratory. However, some of the genes identified by STM (e.g. the O-antigen biosynthesis genes) are expected to be expressed significantly in the laboratory. Another difference is that IVET identifies promoter fusions, whereas STM identifies null mutants. Therefore, STM cannot identify essential genes. IVET can

identify this class of genes, provided that their expression in the laboratory is quite low and less than during infection of the host. An example is provided by the *rpoE* gene, identified by IVET, and later shown to be an essential *Y. enterocolitica* gene (Heusipp *et al.*, 2003). Therefore, it is possible that some of the other genes identified by IVET are also essential, which would explain their absence amongst the genes found by STM. However, it is clear that not all of the IVET genes are essential, and in the cases where null mutants were shown to reduce bacterial load in the host, one would expect these genes to be identified by STM. Of course, there are some technical considerations that may also explain the lack of overlap. For example, mutants that could not grow on minimal media were excluded from the STM screen. The transposon insertion library (STM) and operon fusion libraries (IVET) may not have been completely random, and were certainly not comprehensive. It is possible for some of the IVET operon fusion integrants to be polar within operons that might be required for virulence, preventing the survival of these strains in the animal.

Regardless of the reasons behind the lack of overlap, both IVET and STM have been successful in identifying novel *Yersinia* virulence genes. Many of the genes identified are involved in metabolic functions, and this is beginning to give us a picture of the physiology of *Yersinia* cells during host infection. Whilst these may not be so-called "classical virulence factors", their identification and future characterization is of great importance and interest. After all, the most successful antibacterial agents to date are antibiotics, which do not specifically target virulence factors.

Some of the genes identified are known or thought to encode proteins involved in stress responses (e.g. *acrA*, *clpX*, *dnaJ*, *pspC* and *rpoE*). This apparently points to the importance of the ability of the infecting organism to respond to the environmental changes encountered in the host, such as changes in temperature, pH and osmolarity. The RpoE and PspC proteins are apparently involved in separate extracytoplasmic stress responses, which might be triggered by the production of virulence proteins located in the cell envelope, or that must pass through the cell envelope during their secretion. Indeed, induction of the Psp response is triggered by an essential component of the Ysc type III secretion system (Darwin and Miller, 2001), and by other secretin proteins that play a role in virulence (Maxson and Darwin, 2004). It would be interesting to know whether specific virulence proteins also induce the RpoE response of *Yersinia* species.

The results of the various *Yersinia* IVET and STM screens also demonstrate the importance of using different methods, and different infection models, to identify virulence genes. With the availability of annotated genome sequence information, the characterization of loci identified by these techniques can proceed much more rapidly than previously. Therefore, we have probably not seen the last of the application of these techniques to the study of *Yersinia* pathogenesis.

A brief word about subtractive hybridization

It seems appropriate here to briefly discuss another genome-wide approach that has recently been used to

identify *Yersinia* virulence genes, even though it is not an *in vivo* technology. This method, known as suppressive subtractive hybridization, was used to identify a novel chromosomal locus that was unique to the a *Y. enterocolitica* serotype O:8 biotype 1B strain (Iwobi *et al.*, 2003).

Total genomic DNA of a non-pathogenic *Y. enterocolitica* biotype 1A strain was subtracted from the genome of a highly pathogenic *Y. enterocolitica* serotype O:8 biotype 1B strain. The success of the technique was validated by PCR analysis of the resulting clones, which led to the identification of the *ail* (attachment and invasion locus) and *inv* (invasin) genes, and also several genes from the high pathogenicity island. Therefore, 200 subtracted clones were analyzed, and they generally fell into three categories: sequences similar to known genes from *Yersinia* or other species; sequences similar to phages and mobile genetic elements; sequences with no similarity to the databases. The authors focused on one of the sequences that was homologous to the *epsE* gene of *V. cholerae*, which encodes part of a type II secretion system.

Further analysis indicated that the *epsE* homologue is part of a cluster of genes (named the *yts1* cluster), which is homologous to loci from various other species that encode type II secretion systems. The system shared the most similarity with the Eps system of *V. cholerae*, which exports a protease, a chitinase, and the most important *V. cholerae* virulence factor, cholera toxin. Southern hybridization analysis of various *Yersinia* strains confirmed that the Yts1D secretion system is only present in the highly pathogenic *Y. enterocolitica* strains, and is also not found in any other *Yersinia* species. Interestingly, analysis of the *Y. enterocolitica* strain 8081 genome sequence revealed the presence of another locus encoding a type II secretion system (Yts2). Unlike *yts1*, the *yts2* locus was found to be present in all *Y. enterocolitica* strains tested. However, like *yts1*, the *yts2* locus was not found in *Y. pestis* or *Y. pseudotuberculosis*.

A *yts1E* null mutant was shown to be significantly attenuated in a mouse model of oral infection. 48 hours after infection the bacterial CFU recovered from the liver and spleen were approximately 100-fold lower for the mutant than for the wild type strain. However, the mutant showed only a minor reduction in the CFU recovered from the small intestine and Peyer's patches. Therefore, the Yts1 secretion system appears to be important for later stages of the infection. This could be either the process of dissemination, or the ability to survive in tissues such as the liver and spleen. The authors of this study postulated that the Yts1 secretion system must be responsible for the secretion of one or more virulence factors.

Future prospects

All future studies intended to identify *Yersinia* virulence genes should be greatly impacted by the availability of genome sequence information. At the time this review was written, the genome sequences of three *Y. pestis* strains have been published (Parkhill *et al.*, 2001; Deng *et al.*, 2002; Song *et al.*, 2004), as has that of a *Y. pseudotuberculosis* strain (Chain *et al.*, 2004), and the annotated genome sequence of the highly pathogenic *Y.*

enterocolitica strain 8081 is available online (http://www.sanger.ac.uk/Projects/Y_enterocolitica/). In terms of locating virulence genes, there will be obvious advances facilitated by the availability of this information. First, it will greatly aid any future IVET and STM studies. Only a few base pairs of DNA sequence information will be required to determine the genomic context of an IVET operon fusion, or STM transposon insertion. Not only will this save time and money, but it also increases the ability to interpret limited data. An example of this is demonstrated by one of the mutants from the *Y. enterocolitica* STM screen. Initial sequence analysis revealed no homology to database entries (Darwin and Miller, 1999). However, the analysis of a much larger contig from the genome revealed that the transposon insertion was in the *yspC* gene (see above).

Another advance is that genome sequences provide the ability to use other methods to identify *Yersinia* virulence genes. For example, microarray studies to identify all members of a virulence regulon, or genes that are specifically expressed under conditions found in the host. Bioinformatics will also become important by allowing the identification of candidate virulence genes based on their homology to known virulence factors from other organisms. In the future, it will also be interesting to determine the genome sequences of multiple strains of the same species that differ in their levels of virulence. An example would be to compare the genome sequences of non-pathogenic, low-level pathogenic, and highly pathogenic *Y. enterocolitica* strains.

The screens described in this review have already told us a great deal about the *Yersinia* genes that must be expressed during host infection. Genomics holds the promise of extending these conclusions further. Perhaps it is now reasonable to set a somewhat ambitious long-term goal. This would be to have a comprehensive understanding of all genes that must be expressed during infection in an animal model, for all three of the pathogenic *Yersinia* species. We are already on the way to achieving this goal.

Acknowledgments

Work in my laboratory is supported by National Institutes of Health grant AI52148, and by a grant from the Speaker's Fund for Biomedical Research: Toward the Science of Patient Care, awarded by the City of New York. I thank Virginia Miller and Glenn Young for their permission to cite unpublished data.

References

Achtman, M., Zurth, Z., Morelli, G., Torrea, G., Guisoulet, A., and Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* 96: 14043–14048.

Ben-Gurion, R., and Hertman, I. (1958). Bacteriocin-like material produced by *Pasteurella pestis*. *J. Gen. Microbiol.* 19: 289–297.

Brissette, J.L., Russel, M., Weiner, L., and Model, P. (1990). Phage shock protein, a stress protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 87: 862–866.

Carniel, E. (2001). The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes. Infect.* 3: 561–569.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., Brubaker, R.R., Fowler, J., Hinnebusch, J., Marceau, M., Medigue, C., Simonet, M., Chenal-Francisque, V., Souza, B., Dacheux, D., Elliott, J.M., Derbise, A., Hauser, L.J., and Garcia, E. (2004). Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* 101: 13826–13831.

Chiang, S., and Mekalanos, J. (1998). Use of signature-tagged transposon mutagenesis to identify *Vibrio cholerae* genes critical for colonization. *Mol. Microbiol.* 27: 797–805.

Darwin, A.J. (2005). The phage-shock-protein-response. *Mol. Microbiol.* doi:10.1111/j.1365-2958.2005.04694.x: In press

Darwin, A.J., and Miller, V.L. (1999). Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. *Mol. Microbiol.* 32: 51–62.

Darwin, A.J., and Miller, V.L. (2001). The *psp* locus of *Yersinia enterocolitica* is required for virulence and for growth *in vitro* when the Ysc type III secretion system is produced. *Mol. Microbiol.* 39: 429–444.

De Las Penas, A., Connolly, L., and Gross, C.A. (1997). σ^F is an essential sigma factor in *Escherichia coli*. *J. Bacteriol.* 179: 6862–6864.

Deng, W., Burland, V., Plunkett, G., 3rd, Boutin, A., Mayhew, G.F., Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C., Fetherston, J.D., Lindler, L.E., Brubaker, R.R., Plano, G.V., Straley, S.C., McDonough, K.A., Nilles, M.L., Matson, J.S., Blattner, F.R., and Perry, R.D. 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* 184: 4601–4611.

Dorrell, N., Martino, M.C., Stabler, A., Ward, S.J., Zhang, Z.W., McColm, A.A., Farthing, M.J., and Wren, B.W. (1999). Characterization of *Helicobacter pylori* PldA, a phospholipase with a role in colonization of the gastric mucosa. *Gastroenterology* 117: 1098–1104.

Fernandez, L., Marquez, I., and Guijarro, J.A. (2004). Identification of specific *in vivo*-induced (*ivi*) genes in *Yersinia ruckeri* and analysis of ruckerbactin, a catecholate siderophore iron acquisition system. *Appl. Environ. Microbiol.* 70: 5199–5207.

Flashner, Y., Mamroud, E., Tidhar, A., Ber, R., Aftalion, M., Gur, D., Lazar, S., Zvi, A., Bino, T., Ariel, N., Velan, B., Shafferman, A., and Cohen, S. (2004). Generation of *Yersinia pestis* attenuated strains by signature-tagged mutagenesis in search of novel vaccine candidates. *Infect. Immun.* 72: 908–915.

Foultier, B., Troisfontaines, P., Vertommen, D., Marenne, M.N., Rider, M., Parsot, C., and Cornelis, G.R. (2003). Identification of substrates from the *Yersinia enterocolitica* 1B Ysa type III secretion system. *Infect. Immun.* 71: 242–253.

Galloway, S.M., and Raetz, C.R. (1990). A mutant of *Escherichia coli* defective in the first step of endotoxin biosynthesis. *J. Biol. Chem.* 265: 6394–6640.

Gort, A.S., and Miller, V.L. (2000). Identification and characterization of *Yersinia enterocolitica* genes

- induced during systemic infection. *Infect. Immun.* 68: 6633–6642.
- Green, R.C., and Darwin, A.J. (2004). PspG, a new member of the *Yersinia enterocolitica* phage shock protein regulon. *J. Bacteriol.* 186: 4910–4920.
- Haller, J.C., Carlson, S., Pederson, K.J., and Pierson, D.E. (2000). A chromosomally encoded type III secretion pathway in *Yersinia enterocolitica* is important in virulence. *Mol. Microbiol.* 36: 1436–1446.
- Handfield, M., and Levesque, R.C. (1999). Strategies for isolation of in vivo expressed genes from bacteria. *FEMS Microbiol. Rev.* 23: 69–91.
- Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269: 400–403.
- Heusipp, G., Schmidt, M.A., and Miller, V.L. (2003). Identification of *rpoE* and *nadB* as host responsive elements of *Yersinia enterocolitica*. *FEMS Microbiol. Lett.* 226: 291–298.
- Heusipp, G., Young, G.M., and Miller, V.L. (2001). HreP, an *in vivo*-expressed protease of *Yersinia enterocolitica*, is a new member of the family of subtilisin/kexin-like proteases. *J. Bacteriol.* 183: 3556–3563.
- Hung, D.L., Raivio, T.L., Jones, C.H., Silhavy, T.J., and Hultgren, S.J. (2001). Cpx signaling pathway monitors biogenesis and affects assembly and expression of P pili. *EMBO J.* 20: 1508–1518.
- Iwobi, A., Heesemann, J., Garcia, E., Igwe, E., Noelting, C., and Rakin, A. (2003). Novel virulence-associated type II secretion system unique to high-pathogenicity *Yersinia enterocolitica*. *Infect. Immun.* 71: 1872–1879.
- Karlyshev, A.V., Oyston, P.C.F., K., W., Clark, G.C., Titball, R.W., Winzeler, E.A., and Wren, B.W. (2001). Application of high-density array-based signature-tagged mutagenesis to discover novel *Yersinia* virulence-associated genes. *Infect. Immun.* 69: 7810–7819.
- Kelley, T.M., Stachula, S.A., Raetz, C.R.H., and Anderson, M.S. (1993). The *firA* gene of *Escherichia coli* encodes UDP-3-(R-3-hydroxymyristol)-glucosamine N-acyltransferase. *J. Biol. Chem.* 268: 19866–19874.
- Kennedy, E. (1996). Membrane-derived oligosaccharides (periplasmic beta-D-glucans) of *Escherichia coli*. In: *Escherichia coli and Salmonella: cellular and molecular biology*. F. C. Neidhardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. Brooks Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger, eds. ASM Press, Washington, D.C. p. 1064–1071.
- Lesic, B., and Carniel, E. (2004). The high-pathogenicity island: a broad-host-range pathogenicity island. In: *Yersinia molecular and cellular biology*. E. Carniel and B. J. Hinnebusch, eds. Horizon Bioscience, Wymondham, U.K. p. 285–306.
- Li, C., and Clarke, S. (1992). A protein methyltransferase specific for altered aspartyl residues is important in *Escherichia coli* stationary-phase survival and heat-shock resistance. *Proc. Natl. Acad. Sci. USA.* 89: 9885–9889.
- Mahan, M.J., Slauch, J.M., and Mekalanos, J.J. (1993). Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* 259: 686–688.
- Maxson, M.E., and Darwin, A.J. (2004). Identification of inducers of the *Yersinia enterocolitica* phage shock protein system and comparison to the regulation of the RpoE and Cpx extracytoplasmic stress responses. *J. Bacteriol.* 186: 4199–4208.
- Mecscas, J. (2002). Use of signature-tagged mutagenesis in pathogenesis studies. *Curr. Opin. Microbiol.* 5: 33–37.
- Mecscas, J., Bilis, I., and Falkow, S. (2001). 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.* 67: 2779–2787.
- Model, P., Jovanovic, G., and Dworkin, J. (1997). The *Escherichia coli* phage-shock-protein (*psp*) operon. *Mol. Microbiol.* 24: 255–261.
- Morris, V.L., Jackson, D.P., Grattan, M., Ainsworth, T., and Cuppels, D.A. (1995). Isolation and sequence analysis of the *Pseudomonas syringae* pv. tomato gene encoding a 2,3-diphosphoglycerate-independent phosphoglyceromutase. *J. Bacteriol.* 177: 1727–1733.
- Mukhopadhyay, P., Williams, J., and Mills, D. (1988). Molecular analysis of a pathogenicity locus in *Pseudomonas syringae* pv. *syringae*. *J. Bacteriol.* 170: 5479–5488.
- Nelson, K.M., Young, G.M., and Miller, V.L. (2001). Identification of a locus involved in systemic dissemination of *Yersinia enterocolitica*. *Infect. Immun.* 69: 6201–6208.
- Orth, K. (2002). Function of the *Yersinia* effector YopJ. *Curr. Opin. Microbiol.* 5: 38–43.
- Pace, J., Hayman, M.J., and Galan, J.E. (1993). Signal transduction and invasion of epithelial cells by *S. typhimurium*. *Cell* 72: 505–514.
- Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrell, B.G. (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413: 523–527.
- Schmiel, D.H., Wagar, E., Karamanou, L., Weeks, D., and Miller, V.L. (1998). Phospholipase A of *Yersinia enterocolitica* contributes to pathogenesis in a mouse model. *Infect. Immun.* 66: 3941–3951.
- Song, Y., Tong, Z., Wang, J., Wang, L., Guo, Z., Han, Y., Zhang, J., Pei, D., Zhou, D., Qin, H., Pang, X., Zhai, J., Li, M., Cui, B., Qi, Z., Jin, L., Dai, R., Chen, F., Li, S., Ye, C., Du, Z., Lin, W., Yu, J., Yang, H., Huang, P., and Yang, R. (2004). Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res.* 11: 179–197.
- Titball, R.W. (1993). Bacterial phospholipases C. *Microbiol. Rev.* 57: 347–366.
- Visick, J.E., Cai, H., and Clarke, S. (1998). The L-isoaspartyl protein repair methyltransferase enhances survival of aging *Escherichia coli* subjected to secondary environmental stresses. *J. Bacteriol.* 180: 2623–2629.

- Vuorio, R., Harkonen, T., Tolvanen, M., and Vaara, M. (1994). The novel hexapeptide motif found in the acetyltransferases LpxA and LpxD of lipid A biosynthesis is conserved in various bacteria. *FEBS Lett.* 337: 289–292.
- Vuorio, R., Hirvas, L., and Vaara, M. (1991). The Ssc protein of enteric bacteria has significant homology to the acyltransferases LpxA of lipid A biosynthesis, and to three acetyltransferases. *FEBS Lett.* 292: 90–94.
- Young, G.M., and Miller, V.L. (1997). Identification of novel chromosomal loci affecting *Yersinia enterocolitica* pathogenesis. *Mol. Microbiol.* 25: 319–328.
- Zhang, L., Radziejewshka-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P., and Skurnik, M. (1997). Molecular and chemical characterization of lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O:8. *Mol. Microbiol.* 23: 63–76.

