Annu. Rev. Microbiol. 2000. 54:641–79

PATHOGENICITY ISLANDS AND THE EVOLUTION OF MICROBES

Jörg Hacker
Institut für Molekulare Infektionsbiologie, Universität Würzburg, D-97070 Würzburg, Germany; e-mail: j.hacker@mail.uni-wuerzburg.de

James B. Kaper
Center for Vaccine Development and Department of Microbiology & Immunology, University of Maryland School of Medicine, Baltimore, Maryland 21201; e-mail: jkaper@umaryland.edu

Key Words gene transfer, virulence, genome structure, plasmid, bacteriophage

■ **Abstract** Virulence factors of pathogenic bacteria (adhesins, toxins, invasins, protein secretion systems, iron uptake systems, and others) may be encoded by particular regions of the prokaryotic genome termed pathogenicity islands. Pathogenicity islands were first described in human pathogens of the species *Escherichia coli*, but have recently been found in the genomes of various pathogens of humans, animals, and plants. Pathogenicity islands comprise large genomic regions [10–200 kilobases (kb) in size] that are present on the genomes of pathogenic strains but absent from the genomes of nonpathogenic members of the same or related species. The finding that the G+C content of pathogenicity islands often differs from that of the rest of the genome, the presence of direct repeats at their ends, the association of pathogenicity islands with transfer RNA genes, the presence of integrase determinants and other mobility loci, and their genetic instability argue for the generation of pathogenicity islands by horizontal gene transfer, a process that is well known to contribute to microbial evolution. In this article we review these and other aspects of pathogenicity islands and discuss the concept that they represent a subclass of genomic islands. Genomic islands are present in the majority of genomes of pathogenic as well as nonpathogenic bacteria and may encode accessory functions which have been previously spread among bacterial populations.

CONTENTS

INTRODUCTION

Over the last few years, the genome sequences of >20 bacterial species have been determined and analyzed. It became obvious that microbial genomes consist primarily of core sequences, with a fairly homogeneous G+C content and codon usage, which encode housekeeping functions and which carry gene clusters with relatively low mutational capacity. These core sequences include ribosomal-RNA-specific genes and genes that encode key metabolic proteins such as ATPases [for a review see Hacker et al (37) and Morschhäuser et al (78)]. In addition to the core genome, sequences were discovered that differ from the rest of the genome in their G+C content and codon usage and that have been acquired via horizontal gene transfer. For example, the overall amount of horizontally transferred DNA in the *Escherichia coli* K-12 genome has been calculated to be ∼17% (63). These regions, which have been designated genomic islands, encode accessory functions such as additional metabolic activities, antibiotic resistance, or properties involved in microbial fitness, symbiosis, or pathogenesis.

It has long been known that important pathogenicity factors can be encoded by mobile genetic elements, which are capable of lateral gene transfer. The presence of genes encoding diphtheria toxin of *Corynebacterium diphtheriae* on a lysogenic bacteriophage was discovered nearly half a century ago, and in subsequent years a variety of toxins have been shown to be encoded on bacteriophages, including Shiga toxin of enterobacteria, cholera toxin of *Vibrio cholerae*, neurotoxins of *Clostridium botulinum*, and the cytotoxin of *Pseudomonas aeruginosa* (11, 110). Other examples of mobile virulence elements include the heat-stable enterotoxin gene of *E. coli*, which is part of a transposon, and genes encoding a *V. cholerae* hemagglutinin, which are located on an integron structure (68). In addition, important virulence-associated genes of gram-negative pathogens (e.g. *Shigella flexneri*, *Salmonella enterica* and *Yersinia* spp.) as well as of gram-positive pathogens (e.g. *Clostridium tetani* and *Enterococcus faecalis*) are located on plasmids (33, 48, 82). Other virulence determinants are located on the chromosome, where they are often associated in so-called virulence blocks or virulence cassettes. In the early 1980s, it was discovered that the presence of chromosomally located

virulence-associated genes of uropathogenic *E. coli* (UPEC) differs among *E. coli* strains (38, 66). This observation was also made for intestinal *E. coli* pathogens and other pathogenic bacteria of different species and genera (69). In 1990, these regions were termed pathogenicity islands (PAIs) (34). Although the presence of PAIs was first described for human pathogens of the family *Enterobacteriaceae*, it has become clear in recent years that animal and plant pathogens may also contain PAIs. The majority of PAIs are located on the chromosome; however, they can also be part of bacterial plasmids and phages (37).

This article describes the structures and functions of the various PAIs known to date and discusses virulence functions encoded on PAIs and their impact on infectious diseases. In addition, the mobility of PAI elements is summarized. Last, but not least, we discuss PAIs as examples of genomic islands present in the majority of bacterial genomes, where they act as landmarks for lateral gene transfer. The significance of genomic islands, including PAIs, for microbial evolution is discussed.

COMMON FEATURES OF PATHOGENICITY ISLANDS

PAIs in pathogenic *E. coli* were the first described. It was soon discovered that pathogenic bacteria of species other than *E. coli*, both gram positive and gram negative, contain in their genomes DNA segments which share many of the below mentioned features of PAIs. A simplified model of a bacterial PAI is shown in Figure 1. PAIs possess most, if not all, of the characteristics listed below (36, 37).

PAIs carry genes encoding one or more virulence factors. They were first described in human pathogens but are also present in plant pathogens such as *Pseudomonas syringae* (50). One should keep in mind that in some pathogenic

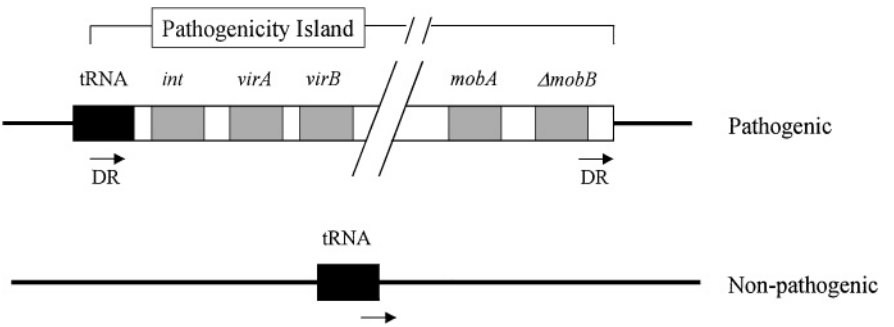

Figure 1 Model of a bacterial pathogenicity island. The *thin bold line* represents regions of the core genome; pathogenicity island-specific sequences are indicated. The *box* represent genes. The *arrows* indicate the presence of direct repeats at the ends of the pathogenicity island. Abbreviations: DR, direct repeats; *int*, integrase gene; *vir*, virulence-associated gene; *mob*, mobility gene; Δmob, pseudo-mobility gene. *mob* genes encode integrases, transposases, or other proteins involved in mobility of the prokaryotic genome.

bacteria virulence factors contribute to pathogenic potential, whereas in other, nonpathogenic bacteria the same factors may be important for survival and replication in particular ecological niches, where they do not create any pathogenic potential (e.g. the iron uptake system in pathogenic *Yersinia* species and nonpathogenic *E. coli*) (88). In nonpathogenic hosts, these DNA segments may act as fitness islands or ecological islands (see section on iron uptake systems below) rather than as PAIs. A summary of important virulence features encoded by PAIs is given in Table 1 and discussed below.

PAIs are present in the genomes of pathogenic organisms but absent from the genomes of nonpathogenic organisms of the same or closely related species. PAIs were first described as chromosomal DNA regions, but the increasing amount of sequence data from extrachromosomal elements supports the view that PAIs may be also part of plasmids or bacteriophage genomes.

PAIs occupy relatively large genomic regions. The majority of PAIs cover DNA regions of $\geq$10–200 kb. Strains of various species, however, may also carry insertions of small pieces of DNA which may encode virulence factors. These DNA regions have been termed pathogenicity islets. From a heuristic point of view, it is useful to distinguish between islands and islets to account for the

TABLE 1 Major virulence features encoded by pathogenicity islands

Virulence feature	Examples
Adherence factors	Diarrheagenic *Escherichia coli* Uropathogenic *Escherichia coli* *Vibrio cholerae* *Listeria* spp.
Toxins	Uropathogenic *Escherichia coli* *Staphylococcus aureus*
Iron uptake systems	Uropathogenic *Escherichia coli* *Shigella flexneri* *Yersinia* spp.
Invasions, modulins, effectors	Diarrheagenic *Escherichia coli* *Salmonella* spp. *Shigella* spp. *Listeria* spp.
Type III secretion systems	Diarrheagenic *Escherichia coli* *Pseudomonas syringae* *Erwinia* spp. *Yersinia* spp. *Salmonella* spp. *Shigella* spp.
Type IV secretion system	*Helicobacter pylori* *Agrobacterium tumefaciens*

apparent differences in size and structure. Some authors, however, prefer the term PAI also for smaller DNA regions (28).

PAIs often consist of DNA regions that differ from the core genome in G+C content and in different codon usage, which may reflect the generation of PAIs by horizontal gene transfer. It should be mentioned here that differences in the G+C contents of PAIs and the core genome will not be observed if the DNAs of the donors and recipients have similar or identical G+C contents.

PAIs are often flanked by small directly repeated (DR) sequences. These sequences may be generated after integration of PAI-specific DNA regions into the host genome via recombination.

PAIs are often associated with transfer RNA (tRNA) genes. tRNA loci often act as integration sites for foreign DNA. The association of PAIs and tRNA loci may therefore reflect the generation of PAIs by horizontal gene transfer.

PAIs often carry cryptic or functional genes encoding mobility factors such as integrases, transposases, and insertion sequence (IS) elements or parts of these elements. PAIs often do not represent homogeneous pieces of DNA but rather are made up of mosaic-like structures which have been generated by a multistep process.

PAIs often represent unstable DNA regions. Deletions of PAIs may occur via the direct repeats (DRs) at their ends or via IS elements or other homologous sequences located on PAIs. Additionally, a few PAIs [e.g. the high PAI (HPI) of *Yersinia pseudotuberculosis*] have the capacity to move from one tRNA site to another (7), and other PAIs may be mobilized and transmitted by bacteriophages (e.g. those of *Staphylococcus aureus* and *V. cholerae*) (56, 64). Particular PAIs represent integrated plasmids, conjugative transposons, or bacteriophages or parts of such elements.

OCCURENCE OF PATHOGENICITY ISLANDS

PAIs occur in the genomes of various human, animal, and plant pathogens. A list of PAIs described up to now (see Table 2) shows that many members of the *Enterobacteriaceae* (e.g. *E. coli, S. flexneri, S. enterica*, and *Yersinia* spp.) cause either intestinal or extraintestinal infections via virulence factors encoded on PAIs. Enterobacteria, as well as *V. cholerae, P. syringae*, and others, show frequent gene transfer via plasmids and bacteriophages. Such extrachromosomal elements indeed seem to represent one source of PAIs (37). Recently it was demonstrated that the genome of the causative agent of Legionnaires' disease, *Legionella pneumophila*, also contains DNA segments which carry type IV secretion-specific genes and which can be considered PAIs (96, 97).

In addition, bacteria that have the capacity to take up DNA from the environment via natural transformation, such as *Helicobacter pylori* and *Neisseria gonorrhoeae*, may carry PAIs (10; JP Dillard & HS Seifert, submitted for publication). It seems, however, that naturally transformable organisms frequently introduce

TABLE 2 Pathogenicity islands (PAIs) and PAI-associated genes of various pathogens[a]

Organism	Description	Functions	Size (kb)	Junction	Associated sequences
E. coli 536 (UPEC)	PAI I_{536}	Hemolysin	70	DR 16 bp	*selC*
E. coli 536 (UPEC)	PAI II_{536}	Hemolysin, P-fimbriae	190	DR 18 bp	*leuX*
E. coli 536 (UPEC)	PAI III_{536}	S-Fimbriae	25		*thrW*
E. coli 536 (UPEC)	PAI IV_{536}	Yersiniabactin synthesis, uptake	45		*asnT*
E. coli J96 (UPEC)	PAI I_{J96}	Hemolysin, P-fimbriae	170		*pheV*
E. coli J196 (UPEC)	PAI II_{J96}	Hemolysin, P-fimbriae, cytotoxic necrotizing factor 1 (CNF1)	110	DR 135 bp	*pheR*
E. coli CFT073 (UPEC)	PAI I_{CFT073}	Hemolysin, P-fimbriae	58	DR 9 bp	*metV*
E. coli K1	*kps* PAI	Capsule	n.d.		*pheV*
E. coli E2348/69 (EPEC)	LEE	Type III secretion, invasion	35		*selC*
E. coli O157:H7 (EHEC)	LEE	Type III secretion, invasion	43		*selC*
E. coli EPEC2	LEE	Type III secretion, invasion	35		*pheU*
E. coli ETEC	Tia-PAI	Invasion	46		*selC*
Y. enterocolitica	HPI	Yersiniabactin synthesis, transport	43	DR 17 bp	*asnT*
Y. pseudotuberculosis	HPI	Yersiniabactin synthesis, transport	36	DR 17 bp	*asnT, U, W*
Y. pestis	HPI (*pgm* locus)	Yersiniabactin synthesis, transport Hemin uptake	102	IS100 DR 17 bp	*asnT*

Yersinia spp.	Yop virulon	Type III secretion, effectors (YOPs)	47		Plasmid
Pathogenic *E. coli*	HPI	Yersiniabactin synthesis, transport	42		*asnT*
S. flexneri	SHI-1	Enterotoxin, protease	51	IS elements	
S. flexneri	SHI-2	Aerobactin synthesis, transport	23–30		*selC*
Shigella spp.	Entry region	Type III secretion, invasion	37		Plasmids
S. typhimurium	SPI-1	Type III secretion, invasion into epithelial cells, apoptosis	40		
S. typhimurium	SPI-2	Type III secretion, invasion into monocytes	40		*valV*
S. typhimurium	SPI-3	Invasion, survival in monocytes	17		*selC*
S. typhimurium	SPI-4	Invasion, survival in monocytes	25		Putative tRNA gene
S. dublin	SPI-5	Enteropathogenesis	7		*serT*
V. cholerae	VPI	TCP-adhesin, regulator	39.5	DR 30 bp	*ssrA*
H. pylori	*cag* Pai	Type IV secretion, *cag*-antigen	40	DR 31 bp	*glr*
D. nodosus	*vap* region	Vap-antigens	12	Dr 19 bp	*serV*
D. nodosus	*vrl* region	Vrl-antigens	27		
N. gonorrhoeae	PAI	Serum resistance, cytotoxin	60–70		
L. pneumophila	*icm/dot* region	Type IV secretion	~20		
	lvh region	Type IV secretion	23		
B. fragilis	BfPAI	Fragilysin, metalloprotease II	6		
B. pertussis	*ptx-ptl* locus	Pertussis toxin			tRNA (Asp)
P. syringae	*hrp* cluster	Type III secretion, effectors	35		Plasmid
X. campestis	*hrp* cluster	Type III secretion, effectors	~20		

(*Continued*)

TABLE 2 (*Continued*)

Organism	Description	Functions	Size (kb)	Junction	Associated sequences
Erwinia spp.	*hrp* cluster	Type III secretion, effectors			
R. solanacearum	*hrp* cluster	Type III secretion, effectors			
A. tumefaciens	T-DNA	Crown gall tumor induction Opine production	20	DR 25 bp	Plasmid
S. aureus	SaPI1	Toxic shock syndrome toxin-1, putative superantigen	15.2	DR 17 bp	
L. monocytogenes	*prf vir* gene cluster	Phospholipases, listeriolysin Metalloprotease, ActA protein regulator	9		
L. ivanovii	*prf vir* gene cluster (LIPI-1)	Phospholipases, listeriolysin Metalloprotease, ActA protein regulator	9		
		Internalins, Sphingomyelinase	18		
	LIPI-2 *inlCD*-region	Internalin CD	4		tRNA Thr
C. difficile	Pathogenicity locus	Tcd toxins	19		

[a]DR, direct repeat; T-DNA, transferred DNA; LEE, Locus of enterocyte effacement; Tia, toxigenic invasion locus A; HPI, high PAI; SHI, *Shigella* PAI; SPI, *Salmonella* PAI; VPI, *Vibrio* PAI; BfPAI, *Bacteriodes fragilis* PAI; SaPI, *Staphylococcus aureus* PAI; LIPI, *Listeria ivanovii* PAI; Tcd, *Clostridium difficile* toxin.

small pieces of DNA into their genomes and that the generation of PAIs represents an exceptional event. Gram-positive bacteria such as *Listeria ivanovii* and *S. aureus* carry "classical" PAIs which exhibit the majority of the features discussed above. Other gram-positive pathogens, such as *Listeria monocytogenes* and *Clostridium difficile*, carry on their genomes so-called virulence or pathogenicity gene clusters, which exhibit a few features of PAIs but differ in many aspects (23, 61, 64). It is obvious from Table 2 that the occurrence of PAIs is not restricted to a particular group of pathogens; rather, PAIs are distributed among all groups of pathogens. In the following sections, virulence factors encoded by PAI-specific genes are described.

VIRULENCE-ASSOCIATED GENES ON PATHOGENICITY ISLANDS

Virulence factors encoded on PAIs represent the entire spectrum of bacterial virulence factors, from adhesins to toxins to host defense avoidance mechanisms. This section briefly reviews the major classes of virulence factors found on PAIs but is by no means an exhaustive summary.

Adhesins

PAIs located in the genomes of various species and pathotypes encode adhesins, which mediate the capacity of microbes to attach to specific eukaryotic receptor molecules. Thus, P fimbriae, which represent important adherence factors of UPEC, are encoded by UPEC-specific PAIs (35). These attachment factors have the capacity to bind to galactose–α1-4–galactose-specific receptor molecules on uroepithelial cells. P-fimbrial genes (*pap* or *prs*) are often linked to gene clusters *hly* and *cfn*, respectively encoding the UPEC-specific toxins α-hemolysin and cytotoxic necrotizing factor 1, a linkage that argues for a strong coevolution of these factors (4). In addition, UPEC, as well as sepsis- or meningitis-causing *E. coli*, have the capacity to produce S fimbriae, which bind to sialic acid-specific receptors on uroepithelial cells and on brain cells (79). S-fimbria-specific genes (*sfa*) are part of a PAI that also carries genes for the iron uptake system *iro*, which was initially found in *Salmonella* spp. (U Dobrindt, G Gottschalk, G Blum-Oehler, J Hacker, unpublished data). Sequence data from PAIs of different UPEC isolates show the presence of additional genes with significant homologies to adhesin gene clusters such as those encoding the *Proteus mirabilis* fimbriae (Pmf), the heat-resistant hemagglutinin of enterotoxigenic *E. coli*, and a saliva-binding protein of *Streptococcus sanguis* (35).

The locus of enterocyte effacement (LEE) PAI of enteropathogenic and enterohemorrhagic *E. coli* (EPEC and EHEC, respectively) encodes an important intestinal adherence factor called intimin [reviewed elsewhere (51, 52, 81)]. Intimin is a 94- to 97-kDa outer-membrane protein encoded by the *eae* gene, which

shows similarity to the *Yersinia* invasin adhesin. The role of intimin in disease has been shown in volunteer and animal studies using isogenic *eae* mutants. EPEC and EHEC interact with intestinal epithelial cells in a characteristic pattern, called attaching and effacing, in which the epithelial cell brush border is effaced and the bacteria intimately adhere to the epithelial cell membrane. Marked cytoskeletal changes, including accumulation of polymerized actin, are seen in the epithelial cell directly beneath the adherent bacteria. Intimin mediates the intimate adherence of the bacteria to the host cell, whereas other factors on the LEE PAI (see below) signal cytoskeletal changes. Intimin binds to the translocated intimin receptor (Tir) protein, which is translocated from the bacterium to the host cell via a type III secretion system encoded on the LEE PAI (58). Tir is also encoded on the LEE PAI and has effector functions in addition to serving as a receptor for intimin (see below).

The *Vibrio* PAI (VPI) of *V. cholerae* (54, 55) encodes the toxin-coregulated pilus (TCP), a type 4 pilus that is an essential intestinal adherence factor for this species. The importance of the TCP as an intestinal colonization factor has been shown in volunteer and animal studies [reviewed by Kaper et al (53)], although the TCP has never been shown to actually mediate adherence to epithelial cells. The TCP definitely mediates interbacterial adherence, and its role in intestinal colonization may be to increase the colonizing mass of bacteria while some other factor directly binds the bacteria to epithelial cells. Downstream of the *tcp* gene cluster on the VPI is the *acf* (accessory colonization factor) gene cluster, whose products play a role in chemotaxis, thereby assisting in intestinal colonization.

Another type 4 pilus of *V. cholerae* is the mannose-sensitive hemagglutinin, which is encoded by a cluster of 16 genes (*msh*) flanked by 7-base-pair (-bp) DR sequences (67). The 16.7-kb *msh* region is inserted between the *yhdA* and *mreB* genes, which are adjacent to each other on the *E. coli* chromosome. Volunteer studies have shown that the mannose-sensitive hemagglutinin is not involved in intestinal colonization or any other aspect of cholera (102), thus indicating that, although this region has features of a PAI, it is not involved in pathogenicity. However, finding that the mannose-sensitive hemagglutinin is essential for formation of biofilms on abiotic surfaces (111) led Marsh & Taylor (67) to call this an environmental persistence island. Another fitness or persistence island, in *V. cholerae* El Tor, encodes an exopolysaccharide (EPS^{ETr}) that is also involved in biofilm formation (117). EPS^{ETr} also mediates the rugose colony type and chlorine resistance. Thus, the fitness/persistence islands encoding the mannose-sensitive hemagglutinin and exopolysaccharide may be essential for maintenence of *V. cholerae* in an environmental reservoir between epidemics.

Secretion Systems

Five distinct mechanisms for extracellular secretion of proteins, known as types I through V, have been described in gram-negative bacteria. Such secretion mechanisms are essential for the extracellular secretion of virulence factors to the surface

of the host cell or their direct translocation into the host cell. Extracellular protein secretion via type II, IV, and V mechanisms requires the *sec* general secretion pathway, whereas proteins secreted via type I and III mechanisms do not require the sec system. Type III and type IV systems are the mechanisms that are most closely associated with PAIs, although types I, II, and V can also be found on PAIs.

Type III Secretion Systems

Type III secretion systems have been found on PAIs in *Salmonella*, *Yersinia*, *Shigella*, EPEC, and EHEC as well as in plant pathogens such as *P. syringae*, *Erwinia* spp., *Xanthomonas campestris*, and *Ralstonia solanacearum* [reviewed by Galán & Collmer (32) and by Hueck (46)]. Type III systems are not found on nonpathogenic members of these species. Some pathogens have two distinct type III systems; for example, *Salmonella* has complete type III systems on both the SPI-1 and SPI-2 PAIs [reviewed by Groisman et al (33); see also Galan (31)], and *Yersinia enterocolitica* has a chromosomally encoded type III system in addition to the plasmid-encoded system (39). For most of these PAIs, genes encoding the type III secretion system and genes encoding the proteins secreted by the type III system comprise the great majority of the genes contained on the PAI. Although the sequence and host cell activities of the secreted proteins differ according to the individual pathogen, the majority of the >20 proteins that comprise the type III secretion apparatus are highly conserved among the different pathogens [reviewed by Hueck (46)]. Greater divergence is seen among the proteins secreted by the type III systems, which are characterized by the lack of a cleavable amino-terminal signal peptide.

The majority of the proteins comprising the type III secretion apparatus are located in the inner membrane, but the functions of many of the type III secretion apparatus proteins are not known. Members of the YscN family (e.g. InvC of *S. typhimurium*, Spa47 of *S. flexneri*, and EscN of EPEC/EHEC) are ATPases that are believed to provide energy for the secretion/assembly process (46). Proteins of the LcrD family (e.g. InvA of *S. typhimurium*, MxiA of *S. flexneri*, and EscV of EPEC/EHEC) and proteins of the YscC family (e.g. InvG of *S. typhimurium*, MxiD of *S. flexneri*, and EscC of EPEC/EHEC) are believed to form channels in the inner and outer membranes, respectively (112). Some type III systems, such as those of *S. typhimurium* (62), *P. syringae* (91), and EPEC (59), have filamentous surface structures that are distinct from flagella and are presumed to function as channels for translocation of effector proteins into the host cell. Type III systems have been proposed to act like a needle and syringe, injecting effector proteins into the host cell.

Thus, PAIs of several mammalian and plant pathogens encode type III secretion systems that are essential for virulence. In many cases, the PAIs are self-contained units containing all genes encoding secreted proteins, the secretion apparatus, the chaperones, etc. Many PAIs also contain genes encoding regulators of the type III secretion system, but in most, if not all, PAIs the PAI-encoded regulators are in turn

regulated by regulatory systems encoded outside the PAI (see below). In at least one case, the LEE PAI of EPEC, the cloned PAI confers all known PAI-associated phenotypes on *E. coli* K-12 (70), but in other cases, such as the closely related LEE of EHEC O157:H7, the cloned PAI does not confer the expected phenotype on K-12 (25). Some proteins secreted by a type III system are encoded outside the PAI but still use the secretion system encoded in the PAI. For example, the SopB and SopE proteins, first identified in *S. enterica* serovar Dublin, are secreted via the type III secretion system encoded on the *Salmonella* SPI-1 PAI that is located at 63 min on the *S. enterica* serovar Typhimurium chromosome. SopB is encoded on the SPI-5 PAI, located at 20 min on the chromosome (114), and SopE is encoded on a bacteriophage inserted at 61 min (76). It is unknown whether the *sopB* and *sopE* genes were originally located on the SPI-1 island and subsequently relocated or they were introduced into *Salmonella* spp. after SPI-1 was introduced and somehow adapted to use the preexisting type III secretion system.

Type IV Secretion Systems

Like type III systems, type IV secretion systems have also been closely associated with PAIs. Type IV systems have been found to be essential for full virulence of a variety of pathogens, including *H. pylori*, *Bordetella pertussis*, *L. pneumophila*, *Agrobacterium tumefaciens*, and *Brucella suis* [reviewed by Burns (9)]. In *H. pylori*, the type IV system encoded on the Cag PAI has recently been shown to mediate the translocation of the CagA protein to host cells (95). CagA plays an important role in the induction of host cellular growth changes induced by *H. pylori*. In *B. pertussis*, the type IV secretion system (also called the Ptl transporter) is essential for secretion of pertussis toxin to mammalian cells (9). A type IV secretion system in *B. suis* was recently shown to be necessary for full virulence of this organism in an in vitro infection model (81a).

The type IV secretion system (which has also been called a type V secretion system, a term that we reserve for autotransporting proteins as discussed below) was first described in the plant pathogen *A. tumefaciens*, wherein it mediates transfer of DNA into plant cells. The type IV system in this species is encoded on the ~200-kb Ti plasmid, which is essential for crown gall tumorigenesis of higher plants [reviewed by Winans et al (113)]. Part of the Ti plasmid, called T-DNA, is transferred into plant cells, where it is integrated into the host genome. The transferred *Agrobacterium* genes are expressed in the plant cell, which causes the infected cells to proliferate, ultimately resulting in crown gall tumors. Approximately 20 *vir* genes are responsible for this transfer of T-DNA, and many of the Vir proteins show sequence similarity to Cag proteins of *H. pylori* and Ptl proteins of *B. pertussis* (e.g. VirB4 is similar to Cag3 and PtlC, VirB7 is similar to CagT and PtlI, etc.). In *A. tumefaciens*, the VirB4, VirB7, VirB9, VirB10, and VirB11 proteins assemble to form the core of the transporter while VirB2 forms a pilus. Little information is available on the functions of the type IV proteins of the mammalian pathogens.

In *L. pneumophila*, a cluster of genes called *dot* (107) or *icm* (96, 97) was shown to be essential for intracellular replication in host cells. The proteins encoded by these genes show sequence similarity to the VirB proteins and are also essential for transfer of plasmid DNA from one bacterial cell to another. The role of this type IV secretion system in the pathogenesis of disease due to *L. pneumophila* is assumed to involve transfer of protein virulence factors, rather than DNA, into eukaryotic cells, but the translocated proteins have not yet been identified. Recently another gene cluster, specific for a second type IV secretion system termed Lvh (for *Legionella vir* homologs), was detected on a putative PAI of the *L. pneumophila* genome; the function of this system, however, has yet to be elucidated (97).

Other Secretion Systems

Type III or type IV secretion systems are invariably associated with PAIs and are nearly always essential for full virulence of a pathogen. Other protein secretion systems of gram-negative bacteria have also been found to be encoded on PAIs, but the association of these other systems with PAIs is not yet as strong as it is for type III or type IV systems. As more genomic sequences become available and the evolution of these secretion systems is further characterized, it is possible that most, if not all, representatives of these specialized protein secretion systems will be recognized as being part of PAIs or other genomic islands.

The type I secretion system was first described in the *E. coli* α-hemolysin, which is encoded on PAIs in most UPEC strains and on the pO157 plasmid of EHEC O157:H7. This secretion system requires three proteins: an inner-membrane ATPase that provides energy for the system (HlyB for *E. coli* hemolysin), a membrane fusion protein that spans the periplasm (HlyD), and the TolC outer-membrane protein [for a review see Dobrindt & Hacker (22)]. Other examples of proteins secreted by type I systems include the *B. pertussis* adenylate cyclase, the *Pasteurella haemolytica* leukotoxin, and proteases of *P. aeruginosa* and *Erwinia chrysanthemi*.

Type II secretion systems are exemplified by the pullulanase secretion system of *Klebsiella oxytoca*, which requires 14 proteins, most of which are located in the inner membrane [reviewed by Lory (65) and by Pugsley et al (85, 86)]. The outer-membrane protein PulD forms a pore in the outer membrane, through which the type II-secreted proteins are believed to pass. PulD is a homolog of the YscC family of proteins, which are involved in type III secretion, and homologs of other components of the pullulanase system are involved in secretion of type 4 pili of *V. cholerae* and other organisms. Type II systems are primarily involved in extracellular secretion of degradative enzymes, although some toxins are also secreted via this mechanism. Proteins secreted via type II systems include cellulase and pectinase enzymes of *Erwinia*, elastase, phospholipase C, and toxin A of *P. aeruginosa*, amylase and protease enzymes of *Aeromonas hydrophila*, and chitinase, protease, and cholera toxin of *V. cholerae*. Although the genes encoding the type II secretion apparatus are usually clustered, the genes encoding proteins secreted by the apparatus are frequently unlinked to these clusters and in one case

may not even be on the same chromosome. In *V. cholerae*, the Eps type II secretion system is encoded on the large chromosome while a protein secreted by this system, the Hap protease, is encoded on the small chromosome (105). Type II systems can be found on mobile genetic elements as well as on the core genomes of nonpathogenic members of the same species. For example, the pO157 virulence plasmid of EHEC encodes an apparently intact type II system (8), but the proteins secreted by this system and the contribution of this system to virulence of EHEC are unknown. A cryptic yet apparently intact type II system is also encoded in the *E. coli* K-12 chromosome (27), but the proteins secreted by this system and the environmental conditions necessary for this system to function are not known.

The type V secretion system encompasses a group of proteins known as autotransporters [reviewed by Henderson et al (42)]. This system has also been called type IV in some publications. Examples of the autotransporters include the immunoglobulin A (IgA) protease of *N. gonorrhoeae*, pertactin of *B. pertussis*, the vacuolating cytotoxin (Vac) of *H. pylori*, SepA of *S. flexneri*, and EspC of EPEC. Rather than a cluster of genes encoding a multicomponent secretion apparatus, the secreted protein and the secretion apparatus are encoded in a single open reading frame. The autotransporters are exported from the cytoplasm via the *sec* pathway with cleavage of an amino-terminal signal peptide. The C-terminal portion of the protein then forms a β-barrel structure which inserts into the outer membrane and serves as a pore for the passage of the mature protein. Some autotransporters, such as pertactin, remain attached to the bacterial cell surface, while other members of this class, such as IgA protease, are cleaved, thereby leaving the C-terminal portion in the outer membrane and releasing the mature protein into the extracellular milieu. Examples of autotransporters which are encoded on PAIs include EspC of EPEC (52) and the IgA protease of *N. gonorrhoeae* (84). The SHI-1 (formerly *she*) PAI of *S. flexneri* contains genes for two different yet related autotransporters, SigA and Pic (formerly ShMu), within a 12-kb region (41, 87).

Invasins, Modulins, and Effectors

Some of the best characterized PAIs provide the host bacterium with the ability to invade epithelial cells and/or modulate host cell activities. At least five PAIs (named SPI-1 through SPI-5; Table 2), encoding various virulence phenotypes, including epithelial cell invasion and macrophage apoptosis (SPI-1), intracellular proliferation and systemic spread (SPI-2), intramacrophage survival (SPI-3 and SPI-4), and intestinal fluid secretion and inflammation (SPI-5), have been described for *Salmonella* [reviewed by Groisman et al (33)]. The best-characterized *Salmonella* PAI, SPI-1, encodes a type III secretion system and several proteins that are translocated into host cells via this type III system [reviewed by Galán (31), Galán & Collmer (32), and Groisman et al (33)]. The SipB and SipC proteins form part of the protein translocation system, allowing translocation of effector proteins into epithelial cells. The SipA protein binds cellular actin and inhibits

depolymerization of actin filaments, thereby promoting localized actin cytoskeleton reorganization. The SopE protein, which is encoded outside SPI-I but secreted via the SPI-1 type III secretion system (see above), initiates the invasion process by activating small GTPases such as Rac and Cdc42, thereby leading to activation of mitogen-activated protein kinase pathways and membrane ruffling. Another translocated protein, SptP, inactivates Rac and Cdc42, thereby restoring normal cell architecture after bacterial internalization (30). The SPI-2 island also encodes a type III secretion system, but the proteins secreted by this system are not as well defined as those secreted by the SPI-1 system. One protein translocated into host cells by the SPI-2 secretion system, SpiC, was recently shown to inhibit cellular trafficking in macrophages (106).

Two PAIs located on plasmids in *Shigella* and *Yersinia* spp. also encode proteins that mediate internalization and other host cell responses. The type III secretion system encoded on the *Shigella* virulence plasmid is essential for invasion of host epithelial cells by this pathogen (73, 82, 104). Several Ipa proteins secreted by this type III system are essential for invasion, although the exact functions of these proteins are unknown. IpaB and IpaC are inserted into host cell membranes (3), and IpaC has been specifically implicated in induction of actin polymerization and filopodia formation (103). The IpaA protein binds to the focal adhesion protein vinculin, resulting in F-actin depolymerization (5). The *Yersinia* Yop virulon protects the organism from macrophages by disrupting the phagocytic and signaling functions of these cells and ultimately inducing apoptosis (15, 16, 48). This region encodes a type III secretion system that translocates a number of effector proteins into host cells. The YopH protein has a tyrosine phosphatase activity, and it inhibits phagocytosis and disrupts focal adhesions of HeLa cells. The YopE protein, for which no enzymatic activity has been described, also inhibits phagocytosis and disrupts the cytoskeleton of HeLa cells. The YpkA protein of *Yersinia pestis* and *Y. pseudotuberculosis*, also known as YopO in *Y. enterocolitica*, is a serine/threonine kinase whose target protein is as yet unidentified. The YopT protein induces a cytotoxic effect in cells, leading to disruption of the actin filaments and alteration of the cell cytoskeleton, while mutants defective for production of YopM exhibit reduced virulence in the mouse by an as-yet-unknown mechanism.

The LEE PAI of EPEC and EHEC encodes a type III secretion system that is essential for formation of the attaching-and-effacing histopathology seen with these pathogens [reviewed by Frankel et al (29) and by Kaper & colleagues (52, 81)]. The EspA, EspB, and EspD proteins are secreted via the type III system and are believed to be part of a protein translocation complex with EspA, forming a filamentous structure on the surface of the bacterium, and with EspB and EspD, perhaps forming a pore in the host cell membrane (29, 109). The Tir protein is translocated into host cell membranes via the type III system and serves as a receptor for the LEE-encoded intimin adhesin (58). Tir may also serve to nucleate host cell actin and to transmit signals to the host cell following intimin-Tir binding.

The human pathogen *L. monocytogenes* and the animal pathogen *L. ivanovii* have the capacity to survive and replicate intracellularly in eukaryotic cells (61).

The internalization of *L. monocytogenes* as well as of *L. ivanovii* is triggered by membrane proteins termed internalins (InlA and InlB). The genes encoding the *L. monocytogenes*-specific large internalins InlA and InlB are located in a particular chromosomal region, but their expression depends on the presence of the PrfA regulator, which is part of the PAI-like Prf virulence gene cluster present in the genomes of both species. In addition to the large internalins InlA and InlB, *Listeria* strains express so-called small internalins (InlC to InlI), which are secreted into the medium and which also contribute to internalization and host cell specificity. The genes encoding the small internalins of *L. ivanovii* are part of two PAIs termed LIPI2 and the *inlCD* gene cluster. Whereas the latter is only 5 kb in size and is linked to a tRNA cluster, LIPI2 is 18 kb and represents an unstable DNA region exclusively located on the *L. ivanovii* genome.

Toxins

Toxin Genes on Mobile Genetic Elements

Many bacterial pathogens harbor plasmids or bacteriophages which encode important toxins. Thus, the heat-labile and heat-stable enterotoxins of enterotoxigenic *E. coli*, as well as pore-forming toxins of various enterobacteria, cytolysins of enterococci, enterotoxins of *S. aureus* with superantigen activity, and neurotoxins of pathogenic clostridia and *Bacillus anthracis*, are plasmid encoded [for a review see Dobrindt & Hacker (22)]. In addition, the cholera toxin genes, the genes encoding diphtheria toxin, the Shiga toxin genes of pathogenic enterobacteria, the *C. botulinum* neurotoxin, and various enterotoxins of *S. aureus* and *Streptococcus pyogenes* with superantigen specificity are located on phages. In addition, unorthodox toxins, factors which modulate host signaling or which may act as effector molecules in bacterial host-cell interactions, are encoded by plasmids and phages. Because so many bacterial toxin genes are located on plasmids and bacteriophages, it is not surprising that PAIs also often carry toxin-encoding genes.

Pore-Forming Toxins

The prototypes of pore-forming PAI-encoded toxins are α-hemolysins of UPEC termed HlyA (22). These toxins are transported by a type I secretion system and have the capacity to lyse erythrocytes and other eukaryotic cells following insertion into the eukaryotic cell membrane. The genes encoding the type I transport system, *hlyB* and *hlyD*, as well as a the *hlyC* gene encoding an HlyA-modifying enzyme, are clustered on PAIs of UPEC and, to a minor extent, on bacterial plasmids. On PAIs, but not on plasmids, the *hly* gene cluster is often colocalized with P-fimbrial genes and loci encoding cytotoxic necrotizing factor 1. The importance of the hemolysin for UPEC is corroborated by the fact that particular strains often carry two PAIs with two *hly* determinants on their genomes. Interestingly, the large virulence plasmid of EHEC carries a *hly* gene cluster similar to the PAI-located *hly* determinant of UPEC (8, 35). Another example of a pore-forming toxin encoded by a PAI or PAI-like structures is listeriolysin O of pathogenic *Listeria* species.

Listeriolysin O is able to lyse erythrocytes and other cells in a cholesterol-dependent manner but is also necessary for lysis of the vacuoles of eukaryotic cells before *L. monocytogenes* spreads through the cytoplasm. The gene encoding listeriolysin O is part of the so-called PrfA virulence gene cluster of *L. monocytogenes* and *L. ivanovii* (also termed LIPI1), which carries virulence genes and the *prf*-specific regulatory region (61). The *prf* locus exhibits some characteristics of PAIs but is considered a chromosomal DNA segment rather than a typical PAI.

Proteases, Lipases, and Enterotoxins

Various bacteria of pathogenic and even nonpathogenic species and subtypes produce enzymes with proteolytic activities, many of which play a role in pathogenicity. The corresponding genes are often located on PAIs or on PAI-like structures of gram-positive bacteria. Thus, the *prf* virulence gene cluster of *L. monocytogenes* and *L. ivanovii* (LIPI1) carries the genes for two phospholipases (*plcA* and *plcB*) and a locus encoding a metalloprotease (*mpl*). In *L. ivanovii*, but not in *L. monocytogenes*, another PAI (termed LIPI2) encodes a sphingomyelinase (SmcL) (61). Furthermore, enterotoxigenic *Bacteroides fragilis* strains produce another metalloprotease, termed fragilysin. This enterotoxin causes fluid accumulation in lamb ligated ileal loops (115). The corresponding gene, *bft*, is inserted into the *B. fragilis* chromosome on a small (6-kb) DNA segment, termed the BfPAI, that appears to be of plasmid origin, suggesting a potential mechanism for introduction into this species (28). In addition, several other proteolytic proteins with enterotoxic activity have been found to be encoded on PAIs. The EspC protein of EPEC, encoded within a 15-kb PAI that is present in some but not all lineages of EPEC, is an autotransporter that increases the short-circuit current in rat jejunal tissue (52). Related proteins with similar functions are the EspP protein, encoded on the EHEC virulence plasmid, and the Pic protease/mucinase of *S. flexneri* SHI-1, which degrades gelatin and mucin. It is interesting that another enterotoxin of *Shigella*, ShET1, which causes fluid accumulation in rabbit ileal loops, is also encoded in the SHI1 PAI of *S. flexneri* by two open reading frames that are on the DNA strand opposite that encodes the Pic protease/mucinase (26, 87).

Toxins with Other Enzymatic Activities

Many toxins have the capacity to modify host cell components such as small G proteins or elongation factors. While some of these toxins (e.g. cholera toxin and diphtheria toxin) are encoded by bacteriophages, others are known to be encoded by PAIs. Thus, the gene encoding cytotoxic necrotizing factor 1 is part of PAIs of UPEC, where it is often located next to *hly* and *pap* gene clusters. cytotoxic necrotizing factor 1 has the capacity to modify the RhoA protein, a small GTPase, by deamidation. Interestingly, the *cnf2* gene, encoding a toxin of similar function, is part of a plasmid of enterotoxigenic *E. coli* (22, 35). Furthermore, the genes encoding the enterotoxins of *C. difficile*, TcdA and TcdB, which exhibit glycosyltransferase activity and also modify GTPases, are part of the pathogenicity

locus, a DNA segment of 19 kb of the *C. difficile* genome which exhibits some, but not all, features of PAIs (23). In addition, the pertussis toxin gene cluster *ptx* of *B. pertussis* has features of PAIs. *ptx* encodes a very complex toxin which ADP-ribosylates small G proteins. Another toxin with enzymatic activity is the Shiga toxin (encoded by *stx*), which specifically cleaves the 28S eukaryotic ribosomal-RNA molecule. Whereas in pathogenic *E. coli* and other enterobacterial species the *stx* genes are part of phage genomes, in *Shigella dysenteriae* the *stx* locus is located on the chromosome. The *stx* region exhibits features of integrated phages which in turn could be considered as PAIs or PAI-like structures (71).

Superantigens

Superantigens have the capacity to bind to the T-cell receptor molecule and in turn activate these cells in a nonspecific manner. Many of these toxins are produced by *S. pyogenes* and *S. aureus*, in which the corresponding genes are often located on bacteriophages or plasmids. The toxic shock syndrome toxin (TSST)-1 toxin of *S. aureus*, however, is part of a 15-kb PAI (SaPI) which represents a defective bacteriophage. The SaPI can be mobilized by helper phages and transferred to other strains (64). Similar PAIs have been found in different *S. aureus* isolates, indicating that the occurrence of SaPIs is a general phenomenon in *S. aureus*.

Iron Uptake Systems

Survival and multiplication of microbes in certain ecological niches depend on the ability of these organisms to scavenge essential nutrients such as iron. For pathogenic bacteria the acquisition of iron is a prerequisite for the infectious process, but nonpathogenic organisms also need sufficient iron ions (74). At least two different strategies are used by bacteria to meet their iron needs: the expression of receptors for iron carriers (e.g. heme, hemoglobin, lactoferrin, and transferrin) and the synthesis and secretion of siderophores, which are low-molecular-weight, high-affinity iron-binding compounds. The gene clusters for two different siderophore systems, yersiniabactin and aerobactin, are part of PAIs, while a hemin uptake system of enterobacteria is part of an islet.

The siderophore system of the phenolate type termed yersiniabactin (Ybt) is encoded by the so-called HPI first described for the pathogenic *Yersinia* spp. *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* (7). The Ybt system, however, is also encoded by nonpathogenic *E. coli*, by defined *E. coli* pathotypes such as enteroaggregative *E. coli* (94), non-O157 EHEC (57), and extraintestinal *E. coli*, as well as by other enterobacteria such as *Citrobacter* spp. and *S. enterica* serotype IIIb (94; T Ölschläger, T Zhang, E Carniel, J Heesemann, W Rabsch, H Tschaepe, H Karch, J Hacker, unpublished data). The HPI consists of a core element, common to all HPI-positive species, comprising 12 yersiniabactin-encoding genes and additional loci at the right end which are specific for particular species and pathotypes. *Y. pestis* and *Y. pseudotuberculosis* also carry hemin storage locus genes, which are located next to the HPI sequences but which are not part of the

island. It is interesting that the HPI is located next to one of three asparagine tRNA loci. In *Y. pseudotuberculosis* the island has the capacity to move from one tRNA locus to the other (7).

Aerobactin has long been known as an important hydroxamate iron uptake system of pathogenic and nonpathogenic *E. coli*, in which the corresponding gene cluster (termed *aer* or *iut*) may be part of large virulence plasmids which also encode the colicin ColV. In *S. flexneri*, however, the *aer* (*iut*) locus is part of a 23- to 30-kb PAI which is located next to the *selC* locus. The island, termed *Shigella* PAI 2 (SHI-2), is unstable, and in species others than *S. flexneri* (e.g. *S. boydii* and *E. coli*) the aerobactin PAI seems to be located at a chromosomal site different from that of *selC* (80, 108). The finding that the *aer* determinant of the *Shigella* PAI is linked to sequences with homology to ColV-specific genes confirms the view that the PAI may be generated following integration of plasmids encoding the aerobactin system as well as colicin biosynthesis. Whereas the two siderophore systems yersiniabactin and aerobactin are encoded by PAI genes, the hemin uptake system *chu* (for *E. coli* hemin uptake) or *shu* (for *Shigella* hemin uptake) is encoded by a 5-kb DNA fragment located at position 78 on the *E. coli* linkage map. This region is considered an islet rather an island, but nevertheless the *chu* (*shu*) gene products seem to be important for the pathogenesis of EHEC and *Shigella* (116).

Iron uptake systems are encoded by pathogenic as well as nonpathogenic members of the same species; however, they are more prevalent in pathogenic strains. This supports the hypothesis that iron uptake systems, as part of genomic islands in nonpathogenic strains, may contribute to the fitness of these organisms as well as their adaptability to particular environments, whereas in pathogenic bacteria these systems contribute to virulence. Accordingly, in nonpathogenic strains such DNA fragments are considered fitness islands while in pathogenic strains they represent PAIs.

Other Virulence-Associated Genes

O Antigen Synthesis

Although most novel pathogens arising by transfer of a PAI are thought to have emerged hundreds, thousands, or millions of years ago, one novel pathogen with a newly acquired (or at least newly recognized) PAI has emerged within the past decade. Historically, *V. cholerae* strains associated with epidemic cholera exclusively possessed the O1 lipopolysaccharide antigen. In 1992, epidemic *V. cholerae* with a different O antigen, O139, emerged in India and Bangladesh. The O139 clone arose from an O1 El Tor strain when a 22-kb region encoding the O1 lipopolysaccharide antigen was replaced by a 35.8-kb region containing the *wbf* genes, which are responsible for synthesis of the O139 polysaccharide [reviewed by Karaolis & Kaper (55)]. Like PAIs, the O139 *wbf* region has a G+C content (average, 42%) lower than that of the genome (47.5%), it contains an IS, and it is inserted at a specific chromosomal site (99). The substitution of the O139 *wbf* genes for the O1 genes had a dramatic effect on the virulence of *V. cholerae* and the

epidemiology of cholera in India and Bangladesh. Before 1992, the majority of cholera cases were in individuals <15 years of age, because substantial immunity is acquired by adulthood in countries in which *V. cholerae* is endemic. When the O139 strain emerged, the majority of cases were in individuals >15 years old whose immunity against O1 strains apparently did not protect against O139 strains (53).

Serum Resistance

PAI-encoded proteins have also been implicated in mediating serum resistance. The *sac-4* gene, encoding serum resistance to *N. gonorrhoeae*, was recently reported to be contained on a PAI that is found preferentially in isolates from persons with disseminated gonococcal infections rather than those from cases of uncomplicated gonorrhea (JP Dillard & HS Seifert, submitted for publication). This island is not present in strain FA1090, the strain for which the *N. gonorrhoeae* genomic sequence was recently obtained. The Pic protein of *S. flexneri* and enteroaggregative *E. coli*, which is encoded on the SHI-1 PAI of *S. flexneri*, confers serum resistance on *E. coli* K-12 strains expressing this protein (41).

Immunoglobulin A Proteases

Proteases of nonenterobacterial species with homology to the degradating proteinase toxins of the Pic family (see section on toxins above) have the capacity to cleave IgA1 molecules. For the enterobacterial IgA-like protease family (Tsh, EspC, EspP, and Pic), an IgA-degrading activity has not been demonstrated so far (41, 42). Other pathogenic organisms, such as *Haemophilus influenzae*, *N. gonorrhoeae*, *Neisseria meningitidis*, and *Streptococcus pneumoniae*, synthesize IgA proteases, while commensal strains of these species do not show this activity. The corresponding genes, however, have not formally been described as PAIs.

Apoptosis

Several PAI-encoded proteins that are secreted via type III secretion systems have been shown to induce apoptosis in host cells. The SipB protein, encoded on the *Salmonella* SPI-1 island, and the IpaB protein, encoded on the *Shigella* virulence plasmid, induce apoptosis in macrophages by binding to caspase-1 (43, 44). The *Y. pseudotuberculosis* YopJ and *Y. enterocolitica* YopP proteins, which are homologous proteins encoded on the *Yersinia* virulence plasmids, also induce apoptosis in macrophages (75, 77). These Yops exhibit a high degree of similarity to AvrRxv from the plant pathogen *X. campestris*. AvrRxv, which mediates a programmed cell death pathway in plant cells, is secreted by the type III secretion system encoded in the *hrp* gene cluster (92), a PAI of plant pathogens.

Capsule Synthesis

Many pathogenic bacteria have the capacity to synthesize capsules; so far, however, the genomic locations of the corresponding genetic determinants have not been carefully described. The capsule synthesis locus of extraintestinal *E. coli*

represents an exception, because in one particular *E. coli* K1 strain the 20-kb segment could be mapped next to the *pheV* tRNA locus. It is still uncertain whether the *kps* gene cluster itself forms a PAI as suggested recently (13). A similar observation was made for the locus encoding the exopolysaccharide PAI of a biofilm-positive subgroup of *S. epidermidis*. The corresponding genes, termed *ica*, are also located on a particular DNA segment of >100 kb which is absent in biofilm-negative strains. Further studies will be neccessary to answer the question of whether the *ica* region can be considered a PAI (118; I Lösner, J Hacker, W Ziebuhr, unpublished data).

Plant Tumorigenesis

The Ti virulence plasmid of *A. tumefaciens* encodes the ability to cause crown gall tumors in higher plants. As noted above, the Ti plasmid encodes a type IV secretion system (encoded by the *virBDE* regions) that transfers one, two, or three fragments of DNA into host plant cells [reviewed by Winans et al (113)]. The transferred DNA, which possesses characteristics more typical of plant DNA than of bacterial DNA, encodes proteins that produce plant hormones such as auxin that cause neoplastic growth of transformed plant cells. Another group of proteins encoded by transferred DNA directs the synthesis by plant cells of opines, which are in turn used by the bacterium for nutrition. The bacterial proteins required for the uptake, degradation, and utilization of opines are encoded by other genes on the Ti plasmid (e.g. *occ*, *mop*, and *aga*).

Genes and Gene Products of Unknown Function

Among the emerging DNA sequence data for partly or entirely sequenced PAIs of various bacteria, many intact open reading frames which may encode putative new virulence factors have been detected. It has been suggested that a functional analysis of these genes and their products will certainly lead to the discovery of new virulence strategies used by PAI-positive bacterial pathogens (35, 83). Additionally, in the genomic islands of the gram-negative pathogen *Dichelobacter nodosus* have been identified the *vap* and *vpl* regions, which exhibit many features of PAIs. The contribution of the *vap* and *vpl* regions and their products to virulence of the bacteria, however, is not yet clear. The DNA sequences of particular open reading frames identified on the *D. nodosus* genomic islands show strong homologies to toxin, adhesin, and enzyme gene clusters of other pathogenic organisms, and it has been suggested that the corresponding products may play a role in the pathogenesis of this important veterinary pathogen (12).

REGULATION OF PATHOGENICITY ISLANDS

PAIs frequently contain genes encoding regulators of virulence factor genes located on the same island. However, various arrangements of regulators and regulatory factors are possible, including PAI-encoded regulators that are specific for PAI

genes, PAI-encoded regulators that also regulate genes located outside the PAI, and regulators encoded outside the PAI that regulate genes encoded on the PAI. The last class includes regulators, present only in pathogens, which regulate primarily virulence factors genes as well as regulators, present in both pathogenic and nonpathogenic members of the species, that may also regulate housekeeping genes in addition to virulence factors. The best-characterized regulatory systems of PAIs feature a regulatory cascade in which PAI-encoded regulators of PAI-encoded virulence genes are themselves regulated by systems encoded outside the PAI.

Two major classes of regulators that are encoded on PAIs or that specifically regulate PAI genes are AraC-like proteins and two-component response regulators. Other classes of regulators include alternate sigma factors and histone-like proteins. A few of the better-characterized PAI regulators are reviewed to illustrate the spectrum of regulatory systems, but for many PAIs, little information concerning regulation is currently available.

The VPI of *V. cholerae* contains at least three regulatory genes, *toxT*, *tcpP*, and *tcpH*. ToxT is a member of the AraC family of transcriptional activators (21) and activates transcription of the *tcp* genes, which code for the major virulence factor encoded on the VPI, TCP. ToxT also regulates the *ctx* genes, encoding cholera toxin, which are located on a filamentous phage inserted outside the VPI. ToxT is itself regulated by TcpP and TcpH, which are cytoplasmic membrane proteins encoded on the VPI (40). Transcription of *toxT* is also regulated by the ToxR and ToxS proteins, which are encoded outside the VPI. ToxR and ToxS are present in nonpathogenic strains of *V. cholerae*, as well as in other *Vibrio* spp., and they regulate the OmpU and OmpT outer-membrane proteins of *V. cholerae*, which are the major porins of this species. Thus, the ToxRS regulon, which controls expression of housekeeping genes in *V. cholerae*, has been adapted by the *toxT* regulator encoded on the VPI. [Expression of *ctx* and *tcp* is also regulated by other factors encoded outside the VPI, such as cAMP receptor protein (CRP) (98)]. It is interesting that TCP and ToxRS are encoded on the larger of the two *V. cholerae* O1 chromosomes, whereas in the classical biotype, *ctx* genes are contained on both the larger and the smaller chromosomes (105), an unusual example of prokaryotic *trans*-chromosomal regulation.

Another well-characterized example of PAI genes regulated by proteins encoded within and outside of the PAI is the SPI-1 island of *Salmonella* [reviewed by Cotter & Miller (17), Groisman et al (33), and Hueck (46)]. The HilA and InvF proteins are encoded on SPI-1 and belong to the OmpR/ToxR and AraC families of transcriptional activators, respectively. HilA regulates InvF, which in turn regulates expression of the *sip* genes, encoding proteins secreted by the type III secretion system, located on SPI-1 (19). HilA also directly regulates expression of the components of the type III system apparatus. The *hilA* gene is itself regulated by proteins encoded both inside and outside SPI-1. The HilC and HilD proteins, encoded on SPI-1, are members of the AraC family of regulators that derepress expression of *hilA* (93). The PhoP and SirA proteins are encoded outside

SPI-1 and also modulate expression of *hilA*. PhoP is the response regulator of the PhoP/PhoQ two-component regulatory system which responds to environmental Mg^{2+} ion concentrations and acts as a negative regulator of HilA. SirA is the response protein of another two-component system for which the cognate sensor protein has not yet been identified. In contrast to the negative effect of PhoP, SirA acts as a positive regulator of HilA expression. Environmental signals such as pH, osmolarity, oxygen, and Mg^{2+} concentration are detected by such global regulatory systems, which then regulate expression of the SPI-1-encoded virulence factors through HilA, supercoiling, or other mechanisms which have yet to be characterized (17, 33).

The plasmid-encoded Yop virulon of pathogenic *Yersinia* spp. is regulated by a global regulator called VirF (LcrF in *Y. pestis* and *Y. pseudotuberculosis*), which belongs to the AraC family [reviewed by Cornelis et al (16)]. VirF directly binds to the promoter regions of the several operons in the Yop virulon and is itself strongly regulated by temperature, being expressed at 37°C but not at 26°C. This temperature regulation involves a chromosomally encoded protein called YmoA which resembles histone-like proteins and acts as a negative regulator of *virF*. An additional level of control may involve feedback inhibition by a repressor(s) (LrcQ in *Y. pseudotuberculosis* and YscM1 and YscM2 in *Y. enterocolitica*) that is normally expelled via the type III secretion system and is similar to the secreted anti-sigma factor involved in regulation of flagellum synthesis. In addition, another regulator of pathogenic yersiniae, YbtA, is encoded by the HPI. YbtA belongs to the AraC class of regulatory elements and seems to be involved in the expression of both HPI-specific genes and determinants located outside the HPI (88).

The LEE PAIs of EPEC and EHEC encode a transcriptional activator of the H-NS family called Ler which controls expression of the type III secretion system and the Tir/intimin receptor/adhesin (72). Ler also controls expression of fimbriae encoded outside the LEE (24) and is itself regulated by integration host factor (IHF) and quorum sensing (100). The quorum-sensing regulatory system is encoded in pathogenic as well as nonpathogenic *E. coli* strains, but the only *E. coli* genes so far known to be regulated by quorum sensing are the LEE genes. EHEC is notable for the low infectious dose required for disease (∼100 organisms), and it has been proposed that the quorum-sensing system of EHEC allows the organism, in the presence of high concentrations of coliforms in the colon, to induce expression of the intimin and Tir proteins necessary for intestinal colonization (100). Regulation of the EPEC LEE is not identical to that of the EHEC LEE, and variations include the presence of an AraC homolog (PerA), encoded on the large virulence plasmid of EPEC, which regulates expression of the EPEC, but not the EHEC, Ler and the transcription of some LEE genes of EHEC, but not EPEC, by the RpoS alternate sigma factor (100). UPEC strains carry fimbrial adhesin determinants that consist of structural genes and two regulatory loci, which in the case of P fimbriae were termed PapB and PapC. The Prf fimbrial operon located on PAI II of strain 536 also encodes such regulators which activate the S-fimbrial determinant (*sfa*) in addition to the P-fimbrial genes. The *sfa* determinant is located on another PAI.

The process in which one group of PAI-located genes is regulated by regulators encoded by another PAI has been termed cross-talk activation (79).

An alternate sigma factor is also involved in regulation of the *hrp* gene cluster in the plant pathogen *P. syringae* which appears to be the plant pathogen equivalent of the PAIs found in mammalian pathogens [reviewed by Hutcheson (47)]. The *hrp* cluster encodes a type III secretion system that is regulated by four proteins encoded on the *hrp* PAI. HrpL is an alternate sigma factor that is essential for transcription of all genes of the *hrp* regulon except *hrpR*, *hrpS*, and *hrpL*. The HrpL sigma factor is similar to σ^{28}, which controls expression of genes involved in biosynthesis of flagella. Transcription of *hrpL* itself is dependent on the σ^{54} alternate sigma factor and is positively regulated by the HrpR and HrpS proteins. The HrpR and HrpS proteins both exhibit similarity to response regulators of two-component systems such as NtrC but they do not appear to be typical two-component regulators, because they lack the N-terminal domain that functions in typical two-component regulators to modulate activity through phosphorylation. A fourth protein, HrpV, negatively regulates genes in the *hrp* regulon by an unknown method, perhaps via HrpR and HrpS. Homologs of the regulatory components HrpL, HrpS, and HrpV have also been implicated in regulation of type III secretion systems on PAIs in *Erwinia* strains that are pathogenic for plants.

The examples presented above represent only a sampling of the various regulators and regulatory mechanisms found in PAIs. Similar regulatory mechanisms are also seen with genes that are not encoded in PAIs. However, one recently described regulatory mechanism, involving translation of rare tRNAs, may be specific for regulation of genes encoded on PAIs. This mechanism, defined by the so-called minor codon hypothesis, is described below in the section on the role of tRNAs.

JUNCTION SITES OF PATHOGENICITY ISLANDS

PAIs represent distinct DNA segments which differ from the core genome by several features (see above). Therefore, the boundary regions of the majority of PAIs have a specific DNA composition, preferentially of DR DNA sequences. As indicated in Figure 1, one of the two PAI junction sites is very often part of the 3′ end of tRNA genes or of similar gene loci encoding the small regulatory RNAs (e.g. *ssrA* for *D. nodosus*) (12) or the gene *glr*, encoding glutamate racemase (e.g. Cag PAI of *H. pylori*) (10). The length of the DRs ranges from 9 bp (in UPEC CFT073) to 135 bp (PAI in UPEC J96 and LEE PAI in particular EHEC strains). In most cases the DRs comprise 16 to 20 bp (35).

The DR DNA sequences were probably generated following insertion of pre-PAI elements such as bacteriophages or plasmids into the core genomes of the host organisms. It should be mentioned here that the 3′ ends of tRNA loci very often act as attachment sites for the integration of bacteriophages. Thus, the strong association between tRNA genes and PAIs may argue for a development of PAIs

from former bacteriophages (89; see also section on mobility genes on PAIs below). In addition, the DR segments may act as target sequences for the action of integrases/excisionases in deletion processes, since many PAIs have the tendency to delete (see section on transfer and deletion of PAIs below). In most cases these deletion processes are RecA independent, and it seems that particular integrases, preferentially those encoded by PAIs, may be involved in the deletion processes. IS elements, in contrast, flank PAIs in very few cases (e.g. the PAI-like element of the *stx* gene in *S. dysenteriae*) (71). IS elements or portions thereof are, however, very often part of PAI segments. This is also true for Rhs sequences and other repeated DNA segments which do not form junction sites for PAIs but often act as targets for partial deletions of PAI-specific sequences (7).

THE ROLE OF tRNA LOCI

About 75% of the PAIs identified so far are associated with tRNA loci (see Table 2). It has been known for many years that tRNA genes may act as landmarks for the integration of foreign DNA, either of plasmids and phages or, in the case of eukaryotes, of retroviruses (45, 89). It is interesting that extrachromosomal genetic elements themselves often carry tRNA genes or parts of them. This is true for plasmids of streptomycetes but also for bacteriophages of enterobacteria, such as the T4 phage or the phage carrying Shiga toxin genes. In addition, in eukaryotes, mitochondrial as well as plastidic DNA molecules harbor tRNA loci. It is therefore suggested that the extrachromosomal elements integrate into the chromosomes of their host organisms by homologous recombination between tRNA genes of the chromosome and their extrachromosomal counterparts, thereby keeping the tRNA loci intact. It seems that the 3′ ends of the tRNA genes may preferentially act as targets for integration. DNA sequence data for many PAIs and their associated tRNA genes show that overlaps of 15–25 nucleotides exist between the 3′ ends of the tRNA loci and PAI-specific sequences (45).

Although in principle many tRNA genes can be used as integration sites for PAIs and other extrachromosomal elements, *in praxi* particular preferences exist. The tRNA genes frequently associated with PAIs (see Table 2) also quite often harbor attachment sites for bacteriophages, indicative of the development of PAIs from former bacteriophage genomes. One of the tRNA loci most frequently associated with PAIs is the *selC* gene, encoding the selenocysteine-specific $tRNA^{sec}$. In enterobacteria the bacteriophage ϕR73 and five different PAI elements can be located next to *selC* (35). Three different *E. coli*-specific PAIs, leading to different *E. coli* pathotypes, as well as PAIs of *S. enterica* and *S. flexneri* are associated with the *selC* locus. In addition, the two identical phenylalanine-specific tRNA genes *pheV* and *pheU* (also termed *pheR*) often act as targets for PAI formation. The idea that PAIs may represent a particular subtype of genomic island is corroborated by the fact that not only three different PAIs but also two genomic islands, encoding an alternative sucrose uptake system in *S. enterica* and the nitrogenase system in

Mesorhizobium loti, are located next to the *pheR* and *pheU* loci (37; see Tables 2 and 3). In addition, tRNA-associated PAIs also have the capacity to be deleted from the genomes of their respective pathogens, leading to a truncation of the tRNA-specific genes. In the case of UPEC strain 536, the tRNA genes *leuX* and *selC*, located next to PAI I and PAI II, became truncated at their 3′ ends after deletion of the two PAIs, which consequently leads to tRNA-specific mutations. It therefore seems that only tRNA genes with another identical locus in the genome (e.g. phenylalanine-specific tRNAs), genes which encode tRNAs with wobble capacity (e.g. *leuX*), or nonessential tRNAs (e.g. *selC*) are frequently used as integration sites for PAIs.

Many authors have suggested that the presence of particular tRNAs that are less frequently used than others (so-called minor tRNAs) may have a modulatory effect on the translational efficiency of particular target genes (the minor codon hypothesis). The evolutionary advantage of extrachromosomal genetic elements carrying additional tRNA genes may be a more effective expression of genes with a larger amount of the corresponding codon located on the respective element. Thus, the genes present on PAI II of UPEC strain 536 that are located next to the *leuX* tRNA have a significantly larger number of *leuX*-specific codons than genes located on the rest of the core genome. Therefore, it was not surprising that *leuX* mutants were less virulent than their *leuX*-positive counterparts, because many genes located on PAI II, such as that encoding α hemolysin, and also other genes were less frequently expressed (90). For type I fimbriae, whose expression

TABLE 3 Examples of genomic islands

Organism	Property	Type of island	Genetic feature	Size (kb)
Mesorhizobium loti	Nitrogen fixation	Symbiosis	Integrated plasmid	500
Pseudomonas putida	Chlorocatechol degradation	Degradation	Integrated plasmid	105
Salmonella senftenberg	Sucrose uptake	Metabolism	Conjugative transposon	100
Staphylococcus aureus	MecA protein	Resistance	Location on chromosome	52
Salmonella typhimurium DT 104	Multiresistance	Resistance	Location on chromosome	14
Various bacteria	Type III secretion Type IV secretion	Secretion	Location on chromosome or plasmid	Variable
Various bacteria	Iron uptake	Fitness	Location on chromosome or plasmid	Variable

is also affected by the presence of *leuX*, it was shown that the translation of the *fimB*-specific positive regulator gene, which harbors an unusually large number of *leuX* codons, is strongly affected by the presence of the *leuX*-specific $tRNA_5^{Leu}$. It is interesting that the *leuX* gene itself is regulated by global regulators such as the alternative sigma factors RpoS and RpoH, providing additional support for the minor codon hypothesis (U Dobrindt & J Hacker, unpublished data). Besides *leuX*, other minor tRNA genes, such as *argU* and *leuZ*, influence the expression of particular virulence-associated genes such as the fimbrial genes of *S. enterica*.

MOBILITY GENES ON PATHOGENICITY ISLANDS

A variety of mobility genes are found on PAIs, particularly genes encoding phage integrases and ISs/transposons. However, the presence and functionality of these mobility genes can differ greatly among PAIs, from genes that encode intact and functional integrases to mutated genes that encode nonfunctional integrases to PAIs that contain no apparent integrase genes. These mobility genes and the protein products they encode can lead to the formation, integration, deletion, and mobility of PAIs as described below in the section on transfer and deletion of PAIs.

Several PAIs contain sequences that show homology to genes encoding integrases of bacteriophage P4 and retronphage ϕ73. P4 integrase sequences are present in the HPI of *Yersinia*, a PAI which is also found in enteroaggregative *E. coli* (94), some Shiga toxin-producing *E. coli* (STEC) serotypes (57), and UPEC (PAI IV_{536}) (94). This integrase appears to be intact and functional in the majority of strains, in contrast to the P4 integrases encoded in PAI II_{536} of UPEC and in *E. coli* K-12, both of which appear to be defective due to the presence of premature stop codons. Integrases that are more closely related to the integrase of ϕ73 are found in PAI I_{536} of UPEC, the SHI-2 PAI of *Shigella* species (108), and in the *vap* (for virulence-associated protein) regions of *D. nodosus* (12, 35). The integrase present in the *V. cholerae* VPI most closely resembles that of the cryptic bacteriophage CP4-57 (53% amino acid identity) but also exhibits substantial homology with the P4 integrase (32% identity) and the *D. nodosus vap*-encoded integrase (34% identity). The substantial homology among integrases from different genera suggests that these sequences could be transferred among a wide range of bacterial species.

The integrases and other bacteriophage proteins encoded on PAIs are usually assumed to have played a role in the original mobilization of the PAI, but there are some phages that appear to have been inserted into the PAI after the original mobilization of the PAI into the chromosome. One such phage is the 933L prophage of the EHEC O157:H7 LEE, which is absent from the closely related LEE of EPEC E2348/69 (83). Similar phages may have nevertheless played a role in the mobilization of other PAIs, because an integrase gene nearly identical to that of 933L is also present in the *Shigella* SHI-2 PAI (108).

Numerous IS elements and transposon sequences can be found in PAIs. In some cases, such elements may have played a role in the original mobilization or subsequent rearrangement of the PAI, or they may have inserted afterward and played little or no role in PAI formation. The insertion of IS*605* into the *cag* PAI of *H. pylori* appears to have played no role in the original mobilization of this PAI but has played a major role in its subsequent evolution (18). The SHI-2 PAI of *S. flexneri*, which encodes the aerobactin iron transport system, contains numerous partial and complete IS elements, including IS*3*, IS*629*, IS*600*, IS*2*, and IS*1* (80, 108). Comparison of SHI-2 islands from *S. flexneri* strains SA100 (serotype 2a) and M90T (serotype 5a) revealed essentially identical insertion sites, aerobactin genes, and IS elements on the left side of the PAI but substantial differences in IS elements and other sequences on the right side of the island. The SHI-2 PAIs in *Shigella sonnei* and *Shigella boydii* have the same insertion site and aerobactin genes but exhibit major differences in the IS elements on their left sides.

In addition to mediating rearrangements of PAIs over long time frames (macroevolution), IS elements can also mediate important changes over very short time frames, as is seen in the phase variation of the polysaccharide intercellular adhesin of *Staphylococcus epidermidis*, which is a microevolutionary process. Alternating insertion and excision of IS*256* into different sites of the *S. epidermidis ica* gene cluster controls expression of the polysaccharide intercellular adhesin, which is a major component of staphylococcal biofilms (118).

TRANSFER AND DELETION OF PATHOGENICITY ISLANDS

The VPI of *V. cholerae* was recently shown to be mobilizable from one strain of *V. cholerae* to another by transduction (56). This island encodes several proteins whose predicted sequences are similar to those of proteins of bacteriophages or eukaryotic viruses as well as a protein that is highly homologous to the *E. coli* cryptic prophage (CP4-57) integrase (60). Furthermore, the VPI is flanked by 20-bp DR sequences that resemble phage *att* sites (56). Culture supernatants of VPI-positive *V. cholerae* strains contained sequences present on the VPI, as shown by polymerase chain reaction, but did not contain sequences of other chromosomal genes located outside the VPI (56). Cell-free phage preparations from a strain marked by insertion of a kanamycin resistance gene into the VPI were found to be capable of transferring the marked VPI into a previously VPI-negative recipient strain. These results suggest that transfer of the VPI via phage transduction could play an important role in the evolution of pathogenic *V. cholerae* strains from nonpathogenic strains.

The staphyloccocal PAIs (SaPIs) encode the TSST, a property which has been previously shown to be transferable. It became clear from studies by Novick & coworkers that the SaPI is part of a defective bacteriophage of 15 kb which has the capacity to be excised and circulated by helper phages such as ϕ13 and ϕ80α (64).

After excision, the islands are transduced to other staphylococcal strains with high frequencies. Thus, the SaPI belongs to the group of mobile PAIs that may have been derived from bacteriophages.

Phage transfer has been implicated in the transfer of other PAIs and PAI-like structures. The genes responsible for the serotype 1a O antigen of *S. flexneri* are encoded on a 5.8-kb region that is flanked by 45-bp *att* sequences and IS elements (1). The overall G+C content of this area (40%) is significantly below the 50% *Shigella* genome content. On one side of this region is a gene whose predicted protein product shows significant homology to integrases of P22 and other bacteriophages. This region is inserted at the *thrW* tRNA gene, and in fact the 3′ end of this gene is identical to the *attR* sequence. One final example is the SopE protein of *Salmonella*, which is secreted via the type III secretion system encoded on the SPI-1 PAI. The SopE protein is not encoded on SPI-I but instead is encoded within a temperate bacteriophage belonging to the P2 family (76).

Certain PAIs do not have the capacity to move from one bacterial strain to another but are, however, able to jump from one site to another in the genome of a bacterial pathogen. The HPIs of pathogenic enterobacteria, encoding the iron uptake system yersiniabactin in *Y. pseudotuberculosis* and *Y. pestis*, carry intact integrase genes and are flanked by DRs of 17 kb (88). The HPIs in *Yersinia* are usually located in the asparagine-specific tRNA gene *asnT*. In particular strains of *Y. pseudotuberculosis*, however, the HPI first detected in *asnT*, over a short time period, has also been found to be located in two other asparagine-specific tRNA genes, *asnW* and *asnV*, indicating a transfer of the element from one target site to another (7). It has been suggested that such mobility events of PAIs may also occur during infections and may represent examples of PAI-induced microevolutionary processes.

In addition, certain PAIs have a tendency for deletion of particular PAI-specific sequences or of the whole genetic element. On one hand, such deletions may occur over longer time intervals in order to optimize the structure of the PAI elements and to reduce the genetic burden by eliminating genes whose products are not further used (see also below and Figure 2). On the other hand, in contrast to the processes of macroevolution, in many organisms PAIs show deletions in a short time range. Such processes of microevolution have been detected for the Cag PAI of *H. pylori*, which has a tendency to decrease in size over short time intervals (10). In addition, the HPI and also the hemin storage locus of *Y. pestis* carry many Rhs elements which may act as landmarks for frequent deletion processes (7). Furthermore, in other organisms, such as UPEC, whole PAI elements can be deleted. In these cases recombinases/integrases located on the corresponding PAI may be involved in the deletion processes (4, 35). For PAI I and PAI II of *E. coli* 536, however, the PAI-encoded recombinases are not intact, thereby suggesting that integrases of other PAIs or encoded by the rest of the genome are probably involved in the deletion processes. It is suggested that the deletion processes may play a role in the adaptation of pathogenic microbes during certain stages of infection. Therefore, genetic flexibility of pathogenic microbes may create selective advantages over

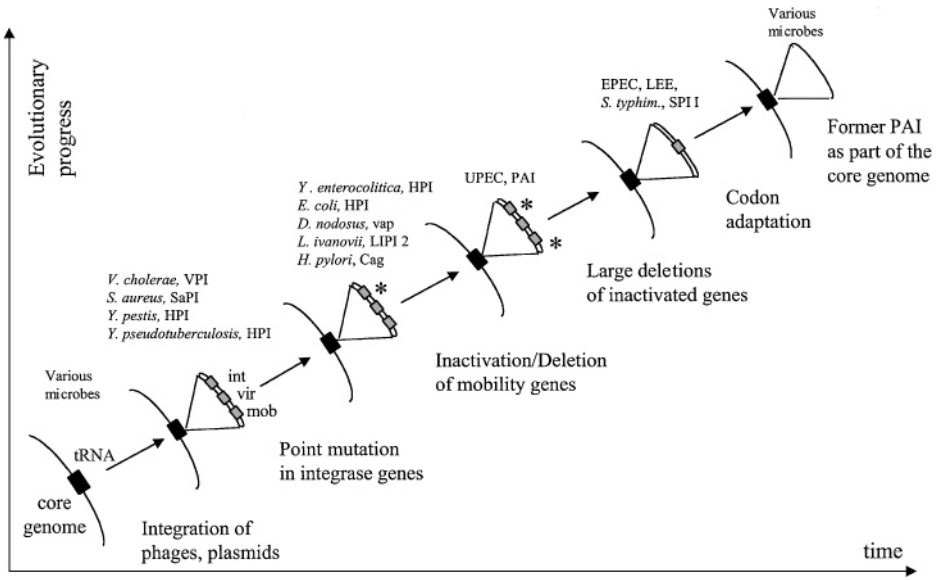

Figure 2 Evolutional stages of pathogenicity island formation. The *vertical line* represents the evolutionary progress; the *horizontal line* represents a time scale. The *thin line* represents parts of the core genome; the *double line* represents pathogenicity island-specific DNA. The *boxes* indicate genes, and the *asterisks* indicate mutations. Abbreviations: tRNA, transfer RNA genes; int, integrase gene; vir, virulence-associated gene; mob, mobility gene (e.g. gene encoding transposase).

other, less-flexible organisms and may finally result in proper replication in host organisms or other ecological niches.

CONCLUSION: Pathogenicity Islands, Genomic Islands, and the Evolution of Microbes

The key processes of Darwinian evolution can be described by the terms genetic variability, phenotype formation, and natural selection. There is no doubt that the permanent generation of new genetic variants represents the main prerequisite for evolution. The term macroevolution describes evolutionary processes which occur within a longer period of time and which lead to the formation of new species or pathotypes (78). In contrast, microevolution, which generates new variants of given species or pathotypes, takes only days or weeks. Both processes, macro- and microevolution, are also involved in the generation and adaptation of pathogenic microbes.

The majority of pathogenic bacteria show evidence of horizontal gene transfer, which together with the generation of point mutations and genetic rearrangements represents one major source for the creation of new genetic variants. The analysis

of the structure of PAIs revealed that these genetic elements have been generated after lateral gene transfer processes. Thus, PAIs contributed to the development of pathogenic variants, preferentially in macroevolution. The ongoing processes of rearrangements, deletions, and transfer of PAIs, however, also have a strong impact on microevolution and adaptation of pathogenic microbes during acute infectious processes.

As indicated in Figure 2, PAIs can be considered genetic fossils owing to their particular stages in the evolutionary process of pathogens. Integrated newly acquired DNAs seem to represent, from the evolutionary point of view, the youngest types of PAIs. The complete or incomplete phages which are identical to *S. aureus* PAIs and the *V. cholerae* VPI encoding the TCP may represent PAIs at the beginning of their career as evolutionary chronometers (56, 64). If a PAI significantly contributes to the fitness and probably also the pathogenic potency of its host organism the mobility genes involved in transfer, deletion, or excision will be inactivated or even deleted. In other words, selection for stability and homing of the formerly acquired DNA takes place. The HPIs of *Yersinia*, the PAIs of UPEC, and the Cag island of *H. pylori* may represent such "middle-aged," adolescent-type PAIs (18, 88). Over time, for example periods of thousands of years, the sizes of successful PAIs tend to be reduced to include just the important genes required for their key functions, and the PAIs become more and more condensed (e.g. the LEE island of EPEC) (52). Thus, the *Salmonella* SPI-1, which was introduced into the genome of the common ancestor of *S. enterica* and *Salmonella bongori* 100 million years ago, is already part of the core genome of both organisms (33). It is our assumption that there exist genetic segments which are former PAI regions but already have been adapted to the core genome of the host and which are no longer detectable as PAIs. In these cases the evolutionary progress has terminated.

Over the past few years, research on PAIs has attracted much attention. However, important questions concerning the development and selection of PAIs remain unanswered. One major question is: Where do the PAI-specific DNA segments come from? From our point of view it is very difficult to answer this question. The overall number of prokaryotes in the soils, fresh waters, and oceans has been estimated to comprise 150,000–200,000 different species; however, only ∼4000 have been described so far. The majority of the bacterial species are not culturable in the laboratory, and therefore analyses of their genomes have been largely impossible. It has been suggested that the enormous amount of genetic information present in the genomes of species not yet described may be a source for many genetic elements, including PAIs. In the case of the *Yersinia* HPI it is suggested that members of the *Pseudomonadae* may be donors for this DNA; however, this has not yet been verified. The further development of microbial ecology will certainly open new avenues to approach the important question of the source of PAI-specific DNA in the future.

Another question is: What are the selective forces for the development of PAIs? Pathogenic organisms as well as nonpathogenic microbes have to overcome environmental restrictions for survival and replication and have to compete successfully

in their ecological niches. These ecological processes, more so than the processes of disease formation, are of evolutionary relevance. The occurrence of infectious diseases does not necessarily contribute to the evolutionary fitness of microbes. It is therefore suggested that from an evolutionary view point, many PAIs represent fitness islands because they were selected in the natural environment of microbes to stimulate their survival and replication. Thus, P fimbriae of UPEC, whose genes are components of PAIs, contribute to adherence of these microbes in the gut (14). Iron uptake systems, of course, are necessary for the survival of microbes in various niches. As the ecological niches and reservoirs of many pathogens (e.g. pathogenic *Shigella*) are not yet known, it is not possible now to fully answer the question on the selective pressures leading to PAI generation.

The third question of interest is: What are the host equivalents of PAIs? The acquisition of new genetic material by the microbes, of course, may generate new phenotypes which may interact with host structures. If the acquisition and generation of PAIs were to generate more-aggressive microbes, they would destroy their hosts and would therefore be unsuccessful in evolutionary terms. On the other hand, the generation of new relevant microbial phenotypes for host interaction should also select for new host structures such as receptor variants or alterations in signal transduction pathways. Such changes in host genes will be important for overcoming hyperpathogenicity and for development of new strategies for an appeasement of hosts and microbes. These host strategies are largely unknown to date, and the study of host-pathogen coevolution at the genetic level will be a major topic in the years to come. The development of DNA chips which represent whole host genomes, including the human genome, will certainly help to address these important questions in the future.

The accumulating sequence data for genomes of pathogenic as well as nonpathogenic bacteria have shown that genetic structures similar to PAIs are also part of the genomes of many nonpathogenic bacteria. These regions were termed genomic islands because they were found to possess many features of PAIs, often differing from the rest of the core genome in G+C content and coding usage, and they often carry mobility genes (integrases and transposases). As with PAIs, genomic islands are often be linked to tRNA genes and may be flanked by repeated DNA at their ends. Last, but not least, genomic islands may represent genome regions of high instability. Many genomic islands, of course, do not contribute to pathogenicity, and their gene products are rather involved in symbiosis, in degradation of xenobiotic compounds, in the generation of new metabolic properties, or in the expression of antibiotic resistance phenotypes (6, 37, 49; for details see Table 3). As already mentioned, iron uptake systems and bacterial secretion pathways can be expressed either by pathogenic or by nonpathogenic bacteria. In both cases they seem to contribute to fitness and adaptation of the strains. It therefore appears that PAIs represent a genomic-island subgroup whose members are, by definition, restricted to pathogenic bacteria. The analysis of PAIs, however, not only opens new insights into the evolution of pathogens but also contributes to the formulation of evolutional principles of prokaryotes in general.

ACKNOWLEDGMENTS

We thank Ute Hentschel (Würzburg) for helpful comments and Claudia Borde (Würzburg) for editorial assistance. We apologize to the many colleagues whose work we could not specifically cite due to space limitations. Our own work related to this topic was supported by the Deutsche Forschungsgemeinschaft (SFB479, Projekt A1) and the Fond der Chemischen Industrie as well as by the National Institutes of Health (grants AI21657, AI19716, and AI41325 to JB Kaper).

LITERATURE CITED

1. Adhikari P, Allison G, Whittle B, Verma NK. 1999. Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the *Shigella flexneri* chromosome. *J. Bacteriol.* 181:4711–18
2. Anderson DM, Schneewind O. 1997. A mRNA signal for the type III secretion of Yop proteins by *Yersinia enterocolitica*. *Science* 278:1140–43
3. Blocker A, Gounon P, Larquet E, Niebuhr K, Cabiaux V, et al. 1999. The tripartite type III secreton of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *J. Cell Biol.* 147:683–93
4. Blum G, Ott M, Lischewski A, Ritter A, Imrich H, et al. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immun.* 62:606–14
5. Bourdet-Sicard R, Rudiger M, Jockusch BM, Gounon P, Sansonetti PJ, Tran VN. 1999. Binding of the shigella protein IpaA to vinculin induces F-actin depolymerization. *EMBO J.* 18:5853–62
6. Briggs CE, Fratamico PM. 1999. Molecular characterization of an antibiotic resistance gene cluster of *Salmonella typhimurium* DT104. *Antimicrob. Agents Chemother.* 43:846–49
7. Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E. 1998. The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn tRNA* genes. *Mol. Microbiol.* 30:965–78
8. Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR. 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* 26:4196–204
9. Burns DL. 1999. Biochemistry of type IV secretion. *Curr. Opin. Microbiol.* 2:25–29
10. Cesini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, et al. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type-I specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA* 93:14648–53
11. Cheetham BF, Katz ME. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.* 18:201–8
12. Cheetham BF, Whittle G, Katz ME. 1999. Are the *vap* regions of *Dichelobacter nodosus* pathogenicity islands? See Ref. 51a, pp. 203–18
13. Cieslewicz M, Vimr E. 1997. Reduced polysialic acid capsule expression in *Escherichia coli* K1 mutants with chromosomal defects in *kpsF*. *Mol. Microbiol.* 26:237–49
14. Connell H, Hedlund M, Agace W, Svanborg C. 1997. Bacterial attachment to uroepithelial cells: mechanisms and consequences. *Adv. Dent. Res.* 11:50–58

15. Cornelis GR, Boland A, Boyd AP, Geuijen C, Iriarte M, et al. 1998. The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol. Mol. Biol. Rev.* 62:1315–52
16. Cornelis GR, Wolf-Watz H. 1997. The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol. Microbiol.* 23:861–67
17. Cotter PA, Miller JF. 1998. In vivo and ex vivo regulation of bacterial virulence gene expression. *Curr. Opin. Microbiol.* 1:17–26
18. Covacci A, Rappuoli R. 1999. *cag*, the pathogenicity island of *Helicobacter pylori*, triggers host response. See Ref. 51a, pp. 189–202
19. Darwin KH, Miller VL. 1999. InvF is required for expression of genes encoding proteins secreted by the SPI1 type III secretion apparatus in *Salmonella typhimurium*. *J. Bacteriol.* 181:4949–54
20. Deleted in proof
21. DiRita VJ. 1992. Co-ordinate expression of virulence genes by ToxR in *Vibrio cholerae*. *Mol. Microbiol.* 6:451–58
22. Dobrindt U, Hacker J. 1999. Plasmids, phages and pathogenicity islands: lessons on the evolution of bacterial toxins. In *The Comprehensive Sourcebook of Bacterial Protein Toxins*, ed. J Alouf, J Freer, pp. 3–23. New York: Academic
23. Eichel-Streiber von C, Braun V. 1999. Virulence-associated mobile elements in bacilli and clostridia. See Ref. 51a, pp. 233–64
24. Elliott SJ, Hutcheson SW, Dubois MS, Mellies JL, Wainwright LA, et al. 1999. Identification of CesT, a chaperone for the type III secretion of Tir in enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 33:1176–89
25. Elliott SJ, Yu J, Kaper JB. 1999. The cloned locus of enterocyte effacement from enterohemorrhagic *Escherichia coli* O157:H7 is unable to confer the attaching and effacing phenotype upon *E. coli* K-12. *Infect. Immun.* 67:4260–63
26. Fasano A, Noriega FR, Maneval DR Jr, Chanasongcram S, Russell R, et al. 1995. Shigella enterotoxin 1: an enterotoxin of *Shigella flexneri* 2a active in rabbit small intestine in vivo and in vitro. *J. Clin. Invest.* 95:2853–61
27. Francetic O, Pugsley AP. 1996. The cryptic general secretory pathway (*gsp*) operon of *Escherichia coli* K-12 encodes functional proteins. *J. Bacteriol.* 178:3544–49
28. Franco AA, Cheng RK, Chung GT, Wu S, Oh HB, Sears CL. 1999. Molecular evolution of the pathogenicity island of enterotoxigenic *Bacteroides fragilis* strains. *J. Bacteriol.* 181:6623–33
29. Frankel G, Phillips AD, Rosenshine I, Dougan G, Kaper JB, Knutton S. 1998. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol. Microbiol.* 30:911–21
30. Fu Y, Galan JE. 1999. A salmonella protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* 401:293–97
31. Galán JE. 1999. Interaction of *Salmonella* with host cells through the centisome 63 type III secretion system. *Curr. Opin. Microbiol.* 2:46–50
32. Galán JE, Collmer A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* 284:1322–28
33. Groisman EA, Blanc-Portard AB, Uchiya K. 1999. Pathogenicity islands and the evolution of *Salmonella* virulence. See Ref. 51a, pp. 127–50
34. Hacker J, Bender L, Ott M, Wingender J, Lund B, et al. 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vivo* and *in vitro* in various extraintestinal *Escherichia coli* isolates. *Microb. Pathog.* 8:213–25
35. Hacker J, Blum-Oehler G, Janke B, Nagy G, Goebel W. 1999. Pathogenicity islands of extraintestinal *Escherichia coli*. See Ref. 51a, pp. 59–76
36. Hacker J, Blum-Oehler G, Mühldorfer I,

Tschäpe H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* 23:1089–97
37. Hacker J, Kaper JB. 1999. The concept of pathogenicity islands. See Ref. 51a, pp. 1–11
38. Hacker J, Knapp S, Goebel W. 1983. Spontaneous deletions and flanking regions of the chromosomal inherited hemolysin determinant of an *Escherichia coli* O6 strain. *J. Bacteriol.* 154:1145–52
39. Haller JC, Carlson SJ, Pierson DE. 1999. A chromosomally-encoded type III secretion apparatus of *Yersinia enterocolitica* affects virulence in the mouse model of infection. *Abstr. ASM Conf. Cell Biol. Approach Microb. Pathog., Portland, OR*, p. 25. Washington, DC: Am. Soc. Microbiol.
40. Häse CC, Mekalanos JJ. 1998. TcpP protein is a positive regulator of virulence gene expression in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* 95:730–34
41. Henderson IR, Czeczulin J, Eslava C, Noriega F, Nataro JP. 1999. Characterization of Pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect. Immun.* 67:5587–96
42. Henderson IR, Navarro-Garcia F, Nataro JP. 1998. The great escape: structure and function of the autotransporter proteins. *Trends Microbiol.* 6:370–78
43. Hersh D, Monack DM, Smith MR, Ghori N, Falkow S, Zychlinsky A. 1999. The *Salmonella* invasin SipB induces macrophage apoptosis by binding to caspase-1. *Proc. Natl. Acad. Sci. USA* 96:2396–401
44. Hilbi H, Moss JE, Hersh D, Chen Y, Arondel J, et al. 1998. *Shigella*-induced apoptosis is dependent on caspase-1 which binds to IpaB. *J. Biol. Chem.* 273:32895–900
45. Hou YM. 1999. Transfer RNAs and pathogenicity islands. *Trends Biochem. Sci.* 24:295–98
46. Hueck CJ. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* 62:379–433
47. Hutcheson SW. 1999. The *hrp* cluster of *Pseudomonas syringae*: a pathogenicity island encoding a type III protein translocation complex? See Ref. 51a, pp. 309–29
48. Iriarte M, Cornelis GR. 1999. The 70-kilobase virulence plasmid of Yersiniae. See Ref. 51a, pp. 91–126
49. Ito T, Katayama Y, Hiramatsu K. 1999. Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-Methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob. Agents Chemother.* 43:1449–58
50. Jackson RW, Athanassopoulos GT, Tsiamis G, Mansfield JW, Sesma A, et al. 1999. Identification of a pathogenicity island, which contains genes for virulence and avirulence, on a large native plasmid in the bean pathogen *Pseudomonas syringae* pathovar phaseolicola. *Proc. Natl. Acad. Sci. USA* 96:10875–80
51. Kaper JB, Gansheroff LJ, Wachtel WR, O'Brien AD. 1998. Intimin-mediated adherence of Shiga toxin-producing *Escherichia coli* and attaching-and-effacing pathogens. In *Escherichia coli O157:H7 and Other Shiga Toxin-Producing E. coli Strains*, ed. JB Kaper, AD O'Brien, pp. 148–56. Washington, DC: Am. Soc. Microbiol.
51a. Kaper JB, Hacker J, eds. 1999. *Pathogenicity Islands and Other Mobile Virulence Elements*. Washington, DC: Am. Soc. Microbiol.
52. Kaper JB, Mellies JL, Nataro JP. 1999. Pathogenicity islands and other mobile genetic elements of diarrheagenic *Escherichia coli*. See Ref. 51a, pp. 33–58
53. Kaper JB, Morris JG Jr, Levine MM. 1995. Cholera. *Clin. Microbiol. Rev.* 8:48–86
54. Karaolis DKR, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR. 1998. A *Vibrio cholerae* pathogenicity

island associated with epidemic and pandemic strains. *Proc. Natl. Acad. Sci. USA* 95:3134–39

55. Karaolis DKR, Kaper JB. 1999. Pathogenicity islands and other mobile virulence elements of *Vibrio cholerae*. See Ref. 51a, pp. 167–87
56. Karaolis DKR, Somara S, Maneval DR Jr, Johnson JA, Kaper JB. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* 399:375–79
57. Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, et al. 1999. A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Escherichia coli* clonal lineages. *Infect. Immun.* 67:5994–6001
58. Kenny B, DeVinney R, Stein M, Reinscheid DJ, Frey EA, Finlay BB. 1997. Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* 91:511–20
59. Knutton S, Rosenshine I, Pallen MJ, Nisan L, Neves BC, et al. 1998. A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells. *EMBO J.* 17:2166–76
60. Kovach ME, Shaffer MD, Peterson KM. 1996. A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio cholerae*. *Microbiology* 142:2165–74
61. Kreft J, Vazquez-Boland JA, Ng E, Goebel W. 1999. Virulence gene clusters and putative pathogenicity islands in Listeriae. See Ref. 51a, pp. 219–32
62. Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, et al. 1998. Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* 280:602–5
63. Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* 95:9413–17
64. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* 29:527–43
65. Lory S. 1998. Secretion of proteins and assembly of bacterial surface organelles: shared pathways of extracellular protein targeting. *Curr. Opin. Microbiol.* 1:27–35
66. Low D, David V, Lark D, Schoolnik G, Falkow S. 1984. Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infections. *Infect. Immun.* 43:353–58
67. Marsh JW, Taylor RK. 1999. Genetic and transcriptional analyses of the *Vibrio cholerae* mannose-sensitive hemagglutinin type 4 pilus gene locus. *J. Bacteriol.* 181:1110–17
68. Mazel D, Dychinco B, Webb VA, Davis J. 1998. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* 280:605–8
69. McDaniel TK, Jarvis KG, Donnersberg MS, Kaper JB. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. USA* 92:1664–68
70. McDaniel TK, Kaper JB. 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol. Microbiol.* 23:399–407
71. McDonough MA, Butterton JR. 1999. Spontaneous tandem amplification and deletion of the Shiga toxin operon in *Shigella dysenteriae* 1. *Mol. Microbiol.* 34:1058–69
72. Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB. 1999. 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.* 33:296–306

73. Ménard R, Sansonetti P, Parsot C, Vasselon T. 1994. Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *S. flexneri*. *Cell* 79:515–25
74. Mietzner TA, Morse SA. 1994. The role of iron-binding proteins in the survival of pathogenic bacteria. *Annu. Rev. Nutr.* 14:471–93
75. Mills SD, Boland A, Sory MP, van der Missen P, Kerbourch C, et al. 1997. *Yersinia enterocolitica* induces apoptosis in macrophages by a process requiring functional type III secretion and translocation mechanisms and involving YopP, presumably acting as an effector protein. *Proc. Natl. Acad. Sci. USA* 94:12638–43
76. Mirold S, Rabsch W, Rohde M, Stender S, Tschäpe H, et al. 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl. Acad. Sci. USA* 96:9845–50
77. Monack DM, Mecsas J, Ghori N, Falkow S. 1997. Yersinia signals macrophages to undergo apoptosis and YopJ is necessary for this cell death. *Proc. Natl. Acad. Sci. USA* 94:10385–90
78. Morschhäuser J, Köhler G, Ziebuhr W, Blum-Oehler G, Dobrindt U, Hacker J. 2000. Evolution of microbial pathogens. *Phil. Trans. R. Soc. London* B 355:695–704
79. Morschhäuser J, Vetter V, Emödy L, Hacker J. 1994. Adhesin regulatory genes within large, unstable DNA regions of pathogenic *Escherichia coli*: cross-talk between different adhesin gene clusters. *Mol. Microbiol.* 11:555–66
80. Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA. 1999. The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol. Microbiol.* 33:74–83
81. Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11:142–201
81a. O'Callaghan D, Cazevielle C, Allardet-Servent A, Boschiroli ML, Bourg G, et al. 1999. A homologue of the *Agrobacterium tumefaciens* VirB and *Bordetella pertussis* Ptl type IV secretion systems is essential for intracellular survival of *Brucella suis*. *Mol. Microbiol.* 33:1210–20
82. Parsot C, Sansonetti PJ. 1999. The virulence plasmid of Shigellae: an archipelago of pathogenicity islands? See Ref. 51a, pp. 151–65
83. Perna NT, Mayhew GF, PósfaI G, Elliott S, Donnenberg MS, et al. 1998. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* 66:3810–17
84. Perrin A, Nassif X, Tinsley C. 1999. Identification of regions of the chromosome of *Neisseria meningitidis* and *Neisseria gonorrhoeae* which are specific to the pathogenic *Neisseria* species. *Infect. Immun.* 67:6119–29
85. Pugsley AP. 1993. The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* 57:50–108
86. Pugsley AP, Francetic O, Possot OM, Sauvonnet N, Hardie KR. 1997. Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in gram-negative bacteria: a review. *Gene* 192:13–19
87. Rajakumar K, Sasakawa C, Adler B. 1997. Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect. Immun.* 65:4606–14
88. Rakin A, Schubert S, Pelludat C, Brem D, Heesemann J. 1999. The high-pathogenicity island of yersiniae. See Ref. 51a, pp. 77–90
89. Reiter W, Palm DP, Yeats S. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res.* 17:1907–14
90. Ritter A, Gally DL, Olsen PB, Dobrindt

U, Friedrich A, et al. 1997. The PAI-associated *leuX* specific $tRNA_5^{Leu}$ affects type 1 fimbriation in pathogenic *Escherichia coli* by control of FimB recombinase expression. *Mol. Microbiol.* 25:871–82

91. Roine E, Wei W, Yuan J, Nurmiaho-Lassila E-L, Kalkkinen N, et al. 1997. Hrp pilus: an *hrp*-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. USA* 94:3459–64
92. Rossier O, Wengelnik K, Hahn K, Bonas U. 1999. The *Xanthomonas* Hrp type III system secretes proteins from plant and mammalian bacterial pathogens. *Proc. Natl. Acad. Sci. USA* 96:9368–73
93. Schechter LM, Damrauer SM, Lee CA. 1999. Two AraC/XylS family members can independently counteract the effect of repressing sequences upstream of the *hil*A promoter. *Mol. Microbiol.* 32:629–42
94. Schubert S, Rakin A, Karch H, Carniel E, Heeseman J. 1998. Prevalence of the high pathogenicity island of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect. Immun.* 66:480–85
95. Segal ED, Cha J, Lo J, Falkow S, Tompkins LS. 1999. Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* 96:14559–64
96. Segal G, Purcell M, Shuman HA. 1998. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc. Natl. Acad. Sci. USA* 95:1669–74
97. Segal G, Russo JJ, Shuman HA. 1999. Relationships between a new type IV secretion system and the *icm/dot* virulence system of *Legionella pneumophila*. *Mol. Microbiol.* 34:799–809
98. Skorupski K, Taylor RK. 1997. Cyclic AMP and its receptor protein negatively regulate the coordinate expression of cholera toxin and toxin-coregulated pilus in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* 94:265–70
99. Sozhamannan S, Deng YK, Li M, Sulakvelidze A, Kaper JB, et al. 1999. Cloning and sequencing of the genes downstream of the *wbf* gene cluster of *Vibrio cholerae* serogroup O139 and analysis of the junction genes in other serogroups. *Infect. Immun.* 67:5033–40
100. Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB. 1999. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 96:15196–201
101. Stein MA, Leung KY, Zwick M, del Portillo FG, Finlay BB. 1996. Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. *Mol. Microbiol.* 20:151–64
102. Tacket CO, Taylor RK, Losonsky G, Lim Y, Nataro JP, et al. 1998. Investigation of the roles of toxin-coregulated pili and mannose-sensitive hemagglutinin pili in the pathogenesis of *Vibrio cholerae* O139 infection. *Infect. Immun.* 66:692–95
103. Tran Van Nhieu G, Caron E, Hall A, Sansonetti PJ. 1999. IpaC induces actin polymerization and filopodia formation during *Shigella* entry into epithelial cells. *EMBO J.* 18:3249–62
104. Tran Van Nhieu G, Sansonetti PJ. 1999. Mechanism of *Shigella* entry into epithelial cells. *Curr. Opin. Microbiol.* 2:51–55
105. Trucksis M, Michalski J, Deng YK, Kaper JB. 1998. The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc. Natl. Acad. Sci. USA* 95:14464–69
106. Uchiya K, Barbieri MA, Funatao K, Shah AH, Stahl PD, Groisman EA. 1999. A

Salmonella virulence protein that inhibits cellular trafficking. *EMBO J.* 18:3924–33
107. Vogel JP, Andrews HL, Wong SK, Isberg RR. 1998. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279:873–76
108. Vokes SA, Torres AG, Reeves SA, Payne SM. 1999. The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. *Mol. Microbiol.* 33:63–73
109. Wachter C, Beinke C, Mattes M, Schmidt MA. 1999. Insertion of EspD into epithelial target cell membranes by infecting enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 31:1695–707
110. Waldor MK, Mekalanos JJ. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272:1910–14
111. Watnick PI, Fullner KJ, Kolter R. 1999. A role for the mannose-sensitive hemagglutinin in biofilm formation by *Vibrio cholerae* El Tor. *J. Bacteriol.* 181:3606–9
112. Wattiau P, Woestyn S, Cornelis G. 1996. Customized secretion chaperones in pathogenic bacteria. *Mol. Microbiol.* 20:255–62
113. Winans SC, Kalogeraki V, Jafri S, Akakura R, Xia Q. 1999. Diverse roles of *Agrobacterium* Ti plasmid-borne genes in the formation and colonization of plant tumors. See Ref. 51a, pp. 289–307
114. Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE. 1998. Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. *Mol. Microbiol.* 29:883–91
115. Wu S, Lim KC, Huang J, Saidi RF, Sears CL. 1998. *Bacteroides fragilis* enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc. Natl. Acad. Sci. USA* 95:14979–84
116. Wyckoff EE, Duncan D, Torres AG, Mills M, Maase K, Payne SM. 1998. Structure of the *Shigella dysenteriae* heaem transport locus and its phylogenetic distribution in enteric bacteria. *Mol. Microbiol.* 28:1139–52
117. Yildiz FH, Schoolnik GK. 1999. *Vibrio cholerae* O1 El Tor: identification of a gene cluster required for the rugose colony type, exopolysaccharide production, chlorine resistance, and biofilm formation. *Proc. Natl. Acad. Sci. USA* 96:4028–33
118. Ziebuhr W, Krimmer V, Rachid S, Lossner I, Götz F, Hacker J. 1999. A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Mol. Microbiol.* 32:345–56

Annual Review of Microbiology
Volume 54, 2000

CONTENTS