Pathogenicity Islands in Bacterial Pathogenesis

Herbert Schmidt* and Michael Hensel*

Institut für Medizinische Mikrobiologie und Hygiene, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Dresden, and Institut für Klinische Mikrobiologie, Immunologie und Hygiene, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

INTRODUCTION
STRUCTURE OF PAI
VIRULENCE FACTORS ENCODED BY PAI
PROTEIN SECRETION SYSTEMS ENCODED BY PAI
Type I Systems
Type II Systems
Type III Systems
Type IV Systems
Type V Systems
REGULATION OF PAI-ENCODED VIRULENCE FUNCTIONS
EVOLUTION AND TRANSFER OF PAI
PAI and Plasmids
Transduction
INTEGRATION SITES OF PAI
PAI OF GRAM-NEGATIVE PATHOGENS
Helicobacter pylori
Pseudomonas aeruginosa
Shigella spp.
ShI-O
ShI-1
ShI-2
ShI-3
Other Shigella islands
Yersinia spp.
HPI
Chromosomally encoded type III secretion systems
Vibrio cholerae
VPI
VPI-2
VSP-I and VSP-II
Salmonella spp.
SPI-1
SPI-2
SPI-3
SPI-4
SPI-5
Major PAI

* Corresponding author. Mailing address for Herbert Schmidt: Institut für Medizinische Mikrobiologie und Hygiene, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Fetscherstraße 74, D-01307 Dresden, Germany. Phone: 49-(0)351-458-6570. Fax: 49-(0)351-4586310. E-mail: herbert.schmidt@mailbox.tu-dresden.de. Mailing address for Michael Hensel: Institut für Klinische Mikrobiologie, Immunologie und Hygiene, Friedrich-Alexander-Universität Erlangen-Nürnberg, Wasserturmstr. 3-5, D-91054 Erlangen, Germany. Phone: 49-(0)9131-85-23640. Fax: 49-(0)9131-85-22531. E-mail: hensel@mikrobio.med.uni-erlangen.de.
## INTRODUCTION

During the last decade, high-throughput techniques have been developed that allow the sequencing of bacterial chromosomes in a short time. To date, 143 bacterial chromosomes have been sequenced completely, and the genome sequences are available at the National Center for Biotechnological Information (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/mier.html) In parallel, bioinformatics underwent a coevolution with the field of genomics, and now we are able to precisely analyze and compare entire chromosomes (e.g., (http://www.tigr.org and http://www.sanger.ac.uk/). Although we are at the beginning of the understanding of bacterial genome structure and architecture, genomic techniques have shown that bacterial DNA is highly dynamic and that the genetic content of bacterial species is in a permanent flux. Even within a species, chromosome sizes may vary between strains or clinical isolates. The genome sizes of nonpathogenic *Escherichia coli* K-12, enterohemorrhagic *E. coli* (EHEC) O157:H7 strain EDL933, and uropathogenic *E. coli* (UPEC) strain CFT073 are 4,639,221, 5,528,445, and 5,231,428 bp, respectively (22, 271, 369).
The process by which the content and organization of genetic information of a species changes over time is known as genome evolution. This process includes four forms of changes: point mutations and gene conversions, rearrangements (e.g., inversion or translocation), deletions, and insertions of foreign DNA (e.g., plasmid integration, transposition). Gene loss and acquisition are genomic changes that can rapidly and radically alter the lifestyle of a bacterium in “quantum leaps” (122). These latter mechanisms seem to be the primary forces by which bacteria genetically adapt to novel environments and by which bacterial populations diverge and form separate, evolutionary distinct species. Acquisition of foreign genes is obviously coupled with gene loss because genome growth is not unlimited. The balance between selective gene acquisition and secondarily imposed gene loss implies that addition of a foreign gene increases the probability of loss of some resident function of lower selective value (207, 260). Mechanisms of horizontal gene flux include mobile genetic elements such as conjugative plasmids, bacteriophages, transposons, insertion elements, and genomic islands, as well as the mechanism of recombination of foreign DNA into host DNA (128, 236).

In this review, we focus on a group of mobile genetic elements whose discovery has influenced and revised our thinking about genomic stability and the species concept in prokaryotes. These elements play a pivotal role in virulence of bacterial pathogens and are also essential for virulence in pathogens of animals and plants. We review a subgroup of genomic islands, the pathogenicity islands (PAI). Since excellent reviews and original papers have already been published on the molecular structure and evolution of these genetic elements (27, 68, 127–129), this review emphasizes the contribution of PAI to the development of disease and to the virulence of bacterial pathogens carrying them.

The concept of PAI was founded in the late 1980s by Jörg Hacker and colleagues in Werner Goebel’s group at the University of Würzburg, Würzburg, Germany, who were investigating the genetic basis of virulence of UPEC strains 536 and J96 (126, 186). The group observed a genetic linkage of determinants encoding P fimbriae, P-related fimbriae, and hemolysins in these strains and could also detect a codeletion of these linked genes (126). Similar DNA segments with more than one linked virulence gene were described earlier and were termed virulence gene blocks in concordance with the names given by other authors (125, 151, 151, 215). However, the observation that a single deletion event results in the loss of two linked virulence gene clusters together with additional DNA segments more than 30 kb apart led to the definition of the epithet “pathogenicity DNA islands” and later on to “pathogenicity islands” (PAI) (26, 126). Hacker and colleagues showed that deletion of a PAI led to a nonpathogenic phenotype of E. coli strain 536, and it has been suggested that such deletions are a genetic mechanism to modulate bacterial virulence. In a later study, the size and genetic structure of these PAI found in E. coli strain 536 were investigated in detail (126, 131, 187). The regions carrying genes for hemolysin production (hly) and P-related fimbriae, termed PAI I and II, were mapped at centisomes 82 and 97 in the E. coli chromosome. It was also shown that the tRNA loci selC and leuX are located at...
TABLE 1. Common features of PAI

<table>
<thead>
<tr>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of virulence genes</td>
</tr>
<tr>
<td>Specific presence in pathogens, absence in benign relatives</td>
</tr>
<tr>
<td>Large distinct chromosomal regions (10 to 200 kb)</td>
</tr>
<tr>
<td>Characteristic base composition different from core genome</td>
</tr>
<tr>
<td>Insertion of PAI adjacent to tRNA genes</td>
</tr>
<tr>
<td>Frequent association with mobile genetic elements, i.e., presence of:</td>
</tr>
<tr>
<td>DR</td>
</tr>
<tr>
<td>Cryptic or functional integrase or transposase</td>
</tr>
<tr>
<td>IS elements</td>
</tr>
<tr>
<td>Chromosomally integrated conjugative transposons, plasmids, and phages</td>
</tr>
<tr>
<td>Genetic instability (if functional mobility elements are present)</td>
</tr>
<tr>
<td>Mosaic structures of several acquisitions</td>
</tr>
</tbody>
</table>

the junction to the chromosome and that direct repeats of 16 and 18 nucleotides flank these PAI. After this initial discovery of PAI, these genetic structures have been found increasingly in other E. coli groups and also in other bacterial species (129).

STRUCTURE OF PAI

Genetic features of PAI (127) are outlined below and summarized in Fig. 1 and Table 1.

(i) PAI carry one or more virulence genes; genomic elements with characteristics similar to PAI but lacking virulence genes are referred to as genomic or metabolic islands.

(ii) PAI are present in the genomes of a pathogenic bacterium but absent from the genomes of a nonpathogenic representative of the same species or a closely related species.

(iii) PAI occupy relatively large genomic regions. The majority of PAI are in the range of 10 to 200 kb.

(iv) PAI often differ from the core genome in their base composition and also show a different codon usage. The base composition is expressed as percentage of guanine and cytosine (G+C) bases, and the average G+C content of bacterial DNA can range from 25 to 75%. Most pathogenic bacterial species have G+C contents between 40 and 60%. The reasons for that variation are not known, but the conservation of a genus- or species-specific base composition is a remarkable feature of bacteria. It is considered that the horizontally acquired PAI still has the base composition of the donor species. On the other hand, it is also observed that the base composition of horizontally acquired DNA will gravitate to the base composition of the recipient’s genome during evolution. Thus it is difficult to explain why “ancient” PAI still show a different base composition. Further factors such as DNA topology or specific codon usage of the virulence genes in PAI may also account for the maintenance of the divergent base composition.

(v) PAI are frequently located adjacent to tRNA genes. This observation gave rise to the hypothesis that tRNA genes serve as anchor points for insertion of foreign DNA that has been acquired by horizontal gene transfer. The frequent insertion at tRNA loci may be explained by the observation that genes encoding tRNAs are highly conserved between various bacterial species. After acquisition by horizontal gene transfer, a DNA fragment that contains a tRNA gene can insert into the recipients genome by recombination between the tRNA genes. The second observation is that certain bacteriophages use tRNA genes as specific insertion points in the host genome. tRNA genes may represent specific anchor points for the integration of foreign DNA.

(vi) PAI are frequently associated with mobile genetic elements. They are often flanked by direct repeats (DR). DR are defined as DNA sequences of 16 to 20 bp (up to 130 bp) with a perfect or nearly perfect sequence repetition. DR might have served as recognition sites for the integration of bacteriophages, and their integration resulted in the duplication of the DR. Furthermore, DR act as recognition sequences for enzymes involved in excision of mobile genetic elements, thus contributing to the instability of a PAI flanked by DR. Deletion of a PAI is probably promoted by the same mechanisms that contribute to the loss of antibiotic resistance factors in the absence of selective pressure. In both situations, the deletion results in a reduction in genome size leading to a reduced generation time that is of advantage in competition with other microbes. PAI often carry cryptic or even functional mobility genes such as integrases or transposases. Integrase, which may have been derived from lysogenic bacteriophages, mediate the integration of the phage genome into the genome of the host bacteria, as well as the excision needed to enter a lytic cycle. Such genes are still functional in certain PAI, and the encoded proteins can mediate the excision of the PAI and its loss. The role of bacteriophages in transfer of PAI is described later in this review. Other PAI contain genes that are similar to integrase and resolvase genes of transposons. These mobile genetic elements can change their location within the chromosome, but transposons can also jump from a chromosomal location into a plasmid and vice versa. Insertion sequence (IS) elements are frequently observed in PAI. Insertion of IS elements can result in the inactivation of genes, but the combination of two or more IS elements can also result in the mobilization of larger portions of DNA. PAI can also represent integrated plasmids, conjugative transposons, bacteriophages or parts of these elements (127).

(vii) PAI often are unstable and delete with distinct frequencies. Virulence functions encoded by certain PAI are lost with a frequency that is higher than the normal rate of mutation. Genetic analyses showed that such mutations are caused not by defects in individual virulence genes within the PAI but, rather, by loss of the large portions of a PAI or even the entire PAI. These mutations can be observed during cultivation of pathogens in vitro, but they are also found in isolates obtained from infected individuals, for example during persistent infections. This indicates that such PAI have an intrinsic genetic instability. The same genetic mechanisms allowing the distribution of PAI by horizontal gene transfer also determine their genetic instability. Several characteristic elements, such as integrases, transposases, and IS elements, have been identified that contribute to mobilization and as well as to instability as described above.

(viii) PAI often represent mosaic-like structures rather than homogeneous segments of horizontally acquired DNA. Some PAI represent an insertion of a single genetic element. Others show a more complex structure, since elements of different origin are present. During evolution, several genetic elements have been acquired independently at different time points and from different hosts. However, these DNA acquisitions integrated at the same position into the chromosome of the recip-
ient bacterial cell. This will result in the accumulation of horizontally acquired elements at a certain location of the chromosome, and the same target structures (e.g. tRNA genes) served repeatedly for the integration of the various elements.

These properties apply to numerous PAI, but with the acquisition an increasing amount of genomic sequence information, it became clear that the genomes of prokaryotes are highly diverse mosaic structures. Besides a core genome, which mostly demonstrates homogeneous G+C content and codon usage, there exists a flexible gene pool that is formed by mobile genetic elements. Although the majority of the genes of the flexible gene pool confer selective advantages to their bacterial recipients, a few represent selfish DNA only promoting their own spread. The latter elements are insertion elements, particularly prophages and restriction/modification systems. PAI are part of that flexible gene pool. Sequencing of entire bacterial genomes revealed a more ubiquitous occurrence of such islands than was previously thought, and this represents a paradigm of more general genetic entities that are present in the genome of many bacteria. Therefore, the designation “pathogenicity islands” has been extended to “genomic islands,” which can encode a wide range of functions. Most genomic islands carry genes useful for the survival and transmission of microbes. In a recent review by Hacker and Carniel (128), the authors propose a model for the development of specialized genomic islands. In this model, a bacterial cell first acquires blocks of genes. Selection processes than may favor the maintenance and development of genomic islands that increase fitness. “Fitness islands” may then specialize as ecological, saprophytic, or symbiosis islands or PAI. This model is based on Darwinian laws and explains the actual situation most satisfactorily.

Genomic islands encode many different functions, which depend largely on the environmental context in which the bacterium grows. PAI, which are the best understood genomic islands known to date, carry clusters of virulence genes whose products contribute to the pathogenicity of the bacterium. In the case of E. coli, such islands have allowed the bacteria to adapt to specific environments and to cause disease (Fig. 2). The division of fitness islands into the different subtypes is based not only on their genetic composition but also on their effects in a specific ecological niche and within a particular organism. This means that the same island may fulfill different functions. For example, the yersiniabactin iron uptake system of Yersinia spp. and several pathogenic enterobacteria is also present in soil bacteria. In the latter group, the yersiniabactin iron uptake system appears as a “fitness islands” allowing life under iron-limiting conditions. If this system is present in a

FIG. 2. Model of the development of PAI of pathogenic E. coli. In this basic model, foreign DNA is acquired by an ancient nonpathogenic E. coli strain, e.g., a normal inhabitant of the gut of vertebrates. In EHEC, a virulence-associated plasmid and at least one Stx-converting phage and several PAI have been acquired and maintained due to the specific adaptation to different environments. Genomic islands, which are present in a specific live environment may specialize and are involved in the development of disease such as (A), diarrhea and hemolytic-uremic syndrome after colonization of the large intestine (A), watery diarrhea after colonization of the small intestine (B), and UTI after survival and colonization of E. coli in the bladder (C). Such events probably have led to the development of specific pathotypes of E. coli, examples of which are EHEC (A), EPEC (B), and UPEC (C). In the model described here, the evolutionary sequence of uptake and incorporation of mobile genetic elements has not been considered. tRNA genes and bacterial phage attachment sites are depicted by grey rectangles with dots and hatched dark grey rectangles, respectively. stx, Shiga toxin gene; OI, O-island; espC, E. coli secreted protease gene.
bacterium that colonizes a host organism, together with other factors this locus will become a PAI (128).

During the last decade, many virulence factors present in PAI have been characterized (see Table 3). Although a number of PAI fit the strict definition of PAI mentioned above, some lack one, two, or more features. In this regard, the designation “islet” (e.g., pathogenicity islet or genomic islet) has been used for virulence gene clusters not fully complying with the PAI definition because of being less than 10 kb (128, 242, 354). Nevertheless, low-G+C content, remnants of bacteriophage genes, or association with mobile genetic elements or tRNA genes may identify them as PAI or as ancestral PAI which have undergone genetic modification and immobilization.

With our current state of knowledge, we would distinguish chromosomal islands from phages and plasmids integrated into the chromosome by the presence of autonomous functional replication origins in the latter group. However, there are examples for transitions between plasmids or phages and PAI (see the discussion of PAI of Staphylococcus aureus), making a precise definition difficult.

### VIRULENCE FACTORS ENCODED BY PAI

Bacterial virulence determinants are predominantly encoded by or associated with mobile genetic elements such as phages, plasmids, insertion elements, or transposons, and a large number of such determinants are located within PAI. The functions of PAI-encoded virulence factors are described in the sections describing the particular PAI. However, such factors can be grouped into larger families, examples of which are shown in Table 2.

Since most pathogenicity factors interact with eukaryotic host cells, they must be exposed either at the surface of the bacterial cell or transported out of the bacterial cell and probably into the eukaryotic cell. To export virulence factors, bacteria have developed at least five different protein secretion systems that are summarized in the following section.

### PROTEIN SECRETION SYSTEMS ENCODED BY PAI

Secretion of proteins is a general requirement for pathogenic and nonpathogenic bacteria. Secreted proteins are required for the assembly of the cell envelope, metabolism, and defense against, and interaction with, host cells during pathogenesis. In gram-positive bacteria, extracellular and surface proteins are secreted by the general secretion pathway. In contrast, the presence of an outer membrane in gram-negative bacteria led to the evolution of a remarkable variety of structurally and functionally different secretion systems. The classification of the secretion systems follows a general convention, and the main features of the systems are presented below. For an instructive overview of these protein secretion systems, see reference 348.

#### Type I Systems

Type I secretion systems (T1SS) have a rather simple assembly of an ATP-binding cassette (ABC) transporter protein located within the inner membrane, a periplasmic protein, and an outer membrane protein that forms the secretion pore. The outer membrane proteins are characterized by the presence of 12 β-sheets that assemble into a β-barrel, a pore in the outer membrane. The ABC transporter is dedicated to the transport of a specific substrate protein. However, the outer membrane proteins can interact with different ABC transporters to secrete a variety of target structures. Substrates of T1SS are delivered into the extracellular medium. With respect to pathogenesis, most relevant substrates of T1SS are hemolysins (reviewed in reference 33). An example of a T1SS encoded by a PAI is the paradigmatic hly operon of UPEC, which is responsible for synthesis, activation, and transport of α-hemolysin (see also Uropathogenic E. coli below).

#### Type II Systems

The type II secretion system (T2SS), also referred to as the main terminal branch of the general secretion pathway, repre-
sents the default machinery for protein secretion in pathogenic and nonpathogenic bacterial species. In gram-positive and gram-negative bacteria, a variety of proteins are transported across the cytoplasmic membrane by the Sec system. These substrate proteins are formed as preproteins with a typical N-terminal signal sequence. After transport across the cytoplasmic membrane, this signal sequence is cleaved by a signal protease. In gram-positive bacteria, this transport is sufficient to release proteins into the extracellular medium, but in gram-negative bacteria, T2SS are employed to transport the periplasmic derivatives of the substrate proteins across the outer membrane. T2SS are composed of at least 12 subunits that are located in the inner membrane, the periplasm, and the outer membrane. Oligomers of the subunits in the outer membrane assembled into a pore are also referred to as secretin. Genes encoding the Sec system and the T2SS are located on the core gene set and are not present within PAI. However, a large number of substrate proteins for T2SS are encoded by genes within PAI, with a variety of these proteins being important for pathogenesis (see reference 336 for an overview).

Type III Systems

Type III secretion systems (T3SS) are complex assemblies that require the function of more than 20 genes for their activity. Many of the subunits of T3SS involved in virulence show similarity to the flagellum assembly machinery system. Similar to flagellum systems, the assembly of an organelle in the cell envelope is observed for T3SS in pathogens (23, 197).

Although termed “secretion systems,” the main function of T3SS is not the secretion of proteins into the medium but rather, the translocation across a third membrane, i.e., the membrane of a eukaryotic host cell. Translocation of such effector proteins into eukaryotic host cells is the basis for specific interference with eukaryotic cells functions, resulting in host cell invasion, inactivation of phagocytic cells, apoptosis, and interference with intracellular transport processes. This form of protein translocation requires contact between the pathogen and the target cell. T3SS-dependent translocation can be observed in extracellular pathogens via the cytoplasmic membrane as well as by intracellular pathogens via the phagosomal membrane. Gene clusters encoding T3SS can be found on virulence plasmids, for example in *Yersinia* and *Shigella* spp., as well as in PAI. PAI encoding T3SS include SPI-1 and SPI-2 of *Salmonella enterica* and the locus of enterocyte effacement (LEE) in enteropathogenic *E. coli*. Further reference of the genetic and biochemical features of T3SS can be found in recent reviews (59, 162).

Type IV Systems

Similar to T3SS, type IV secretion systems (T4SS) are able to translocate proteins into a eukaryotic target cell. T4SS also show a complex structure of at least 10 subunits and are similar to conjugation systems for the transfer of DNA (reviewed in reference 54). The best-studied T4SS is the system of *Agrobacterium tumefaciens* that mediates the translocation of a DNA-protein complex into plant cells, a process required for the induction of tumor formation in plants. Related T4SS are also important in a number of human pathogens, such as *Bordetella pertussis*, *Bartonella* spp., *Legionella pneumophila*, *Brucella* spp., and *Helicobacter pylori*. In *H. pylori*, the T4SS is encoded by the *cag* PAI (49), and in other pathogens the genes encoding T4SS are present in large clusters, suggesting their acquisition in the form of a PAI.

Type V Systems

Type V secretion systems (T5SS) are also referred to as autotransporters (for a review, see reference 144). The entire transport system and the substrate protein are synthesized in a form of a single preproprotein. An N-terminal signal sequence directs the secretion of the preproprotein via the Sec system into the periplasm. After proteolytic cleavage of the signal sequence, the transporter domains of preproprotein oligomers form a β-barrel structure in the outer membrane and the passenger domain of the preproprotein passes through the pore formed by the β-barrel. Finally, proteolytic cleavage allows the release of the passenger domain into the extracellular space. There are various passenger domains secreted by T5SS, e.g., the immunoglobulin G proteases and the VacA toxin. Examples of T5SS encoded by PAI are LPA and the EspC PAI of pathogenic *E. coli*, SPI-3 of *Salmonella enterica*, and SHI-1 of *Shigella flexneri*.

**REGULATION OF PAI-ENCODED VIRULENCE FUNCTIONS**

Like other virulence genes, PAI genes are usually not constitutively expressed but respond to environmental signals. PAI are frequently part of complex regulatory networks that include regulators encoded by the PAI itself, regulators encoded by other PAI, and global regulators encoded elsewhere in the chromosome or by plasmids. PAI regulators, in turn, can also be involved in the regulation of genes that are located outside the PAI.

Most frequently, regulators belong to the AraC/XylS family or to the two-component response regulator family. Alternative sigma factors and histone-like proteins are also involved in PAI regulation. Regulatory cascades, in which PAI-encoded regulators of PAI-located virulence genes are modulated by systems encoded outside the PAI, include VPI of *Vibrio cholerae*, SPI-1 and SPI-2 of *S. enterica*, the Yop virulon of pathogenic *Yersinia* spp., and the LEE of enteropathogenic *E. coli* (EPEC) and EHEC. The regulation of SPI-1, SPI-2, and LEE of *E. coli* have been investigated in a number of studies. Although many details are not yet known, a good overview is available (Fig. 3).

As an example, we briefly describe the regulation of invasion genes of *S. enterica* SPI-1. SPI-1 genes are expressed under conditions imposed on the pathogen by the host microenvironment. Such conditions include the oxygen level, osmolarity, bacterial growth phase, pH value, and, as recently described, the presence of short-chained volatile fatty acids (76). Conditions of low oxygen and high osmolarity induce invasiveness, whereas under high-oxygen conditions, the bacteria remain noninvasive. The transduction of these signals may be dependent on the function of the two-component global regulatory systems EnvZ/OmpR, BarA/SirA, PhoPQ, and PhoRB, as well as on FliZ and Hha, all encoded by genes on the core genome.
SPI-1 encodes a number of transcriptional regulators: HilA plays a central role in controlling SPI-1 gene expression, HilD and HilC interact with a DNA sequence upstream of the HilA promoter, presumably displacing a repressor from this site, and InvF controls expression of the genes encoding the substrate proteins of SPI-1. In a current model, SPI-1 regulation involves a cascade of transcriptional activation in which HilD and HilC, HilA, and InvF (Fig. 3A, dark grey bars) act sequentially to activate T3SS genes. First, HilD and HilC bind to several sites within \( P_{hilA} \) and derepress \( hilA \) transcription. Then HilA binds to \( invF \) and \( prgH \) transcription start sites and activates the expression of \( invD \) and \( prgH \). This results in expression of the genes encoding the T3SS (white bars). InvF is also required for expression of \( sptP \), so it is possible that \( sicP, sptP \) may be cotranscribed with the \( sip \) genes. Two other SPI-1 effectors, SigD (SopB), SopE, SopE2, and other unidentified factors are also expressed from InvF/SicA-dependent promoters. Whereas the HilD-HilA-InvF cascade is most plausible, deviations may occur. A number of environmental signals such as oxygen, osmolarity, growth phase, bile salts, and short-chain fatty acids have been described to modulate SPI-1 expression, probably dependent on the function of the component regulatory systems EnvZ-OmpR, BarA-SirA, PhoPQ, and PhoRB as well as \( FliZ \) and \( Hha \) (for reviews, see references 134 and 214).
evolution and transfer of pai

the observation that important virulence factors are present in very similar forms in different bacteria may be explained by horizontal gene transfer. different scenarios can be considered to explain the transfer between bacterial strains and species.

natural transformation

certain bacteria are capable of natural transformation. during certain phases of growth, transport systems are expressed that allow the uptake of free dna from the environment. although the majority of this foreign dna will be degraded, some fragments that harbor “useful” genes are integrated into the genome of the recipient and maintained. it appears possible that this mechanism allows uptake of dna from distantly related species that will be maintained as the selective pressure selects for the newly acquired features.

pai and plasmids

similar clusters of virulence genes are present in pai and on virulence plasmids, indicating that episomal and chromosomal locations are possible for the same gene cluster. it was observed that certain clusters of virulence genes are present in pai of some pathogens but also on virulence plasmids in other bacteria. the t3ss required for invasion of epithelial cells by shigella spp. is encoded by the mxi and spa genes located on a virulence plasmid, and a related gene cluster that is required for the invasiveness of salmonella enterica is located in spi-1 in a chromosomal location. conjugation can allow the transfer of plasmids between bacteria. these plasmids can then replicate a autonomously from the bacterial chromosome, but under certain conditions plasmids may also integrate into the chromosome. conversely, the formation of episomal elements has been reported for certain pai of staphylococcus aureus. thus, plasmids could be another means of transfer of pai between bacteria.

transduction

bacteriophages have been isolated from virtually all bacterial species; even obligate intracellular pathogens such as chlamydia spp. contain specific phages. bacteriophages are able to transfer bacterial virulence genes as passengers in their genomes. the occasional transfer of virulence genes by phages allows the recipient bacteria to colonize new habitats, such as new host organisms or specific anatomic sites. this extension also allows a more efficient spread of the bacteriophages. thus, the transfer of bacterial virulence genes as passengers in the viral genome can also be an evolutionary benefit for the bacteriophage. a well-characterized example of the contribution of bacteriophages to the evolution of bacterial virulence is found in v. cholerae (see “vibrio cholerae” below).

many pai are too large to be transferred as passengers in bacteriophage genomes. for example, gene clusters on pai encoding t3ss or t4ss comprise 25 to 40 kb dna, which is almost equivalent to the total genome size of a bacteriophage. in these cases, other mechanisms are conceivable. certain bacteriophages are capable of generalized transduction. normally, for the replication of the phage within the host bacterium, copies of the phage genome are packaged into phage heads. during replication, the host dna is fragmented. occasionally, the enzymes involved in packaging the phage genome erroneously pack a fragment of the host genome into the phage head. since the resulting particles are still able to infect a new bacterial host, a fragment of the bacterial dna can be transduced. given sufficient sequence similarity, recombination may occur and the transduced fragment is integrated into the genome of the new host.

pai do not occur only in human pathogens; they have also been found in animal and plant pathogens. examples are the hrp islands of pseudomonas syringae and xanthomonas campes-tris, and islands in animal pathogenic salmonellae and staphylococci. they are distributed throughout the bacterial world, and horizontal transfer may be facilitated by plasmids and phages or by bacteria, which are competent for the uptake of free dna by natural transformation.

integration sites of pai

integration of pai into the bacterial chromosome is a sitespecific event. most pai currently known have inserted at the 3’ end of tRNA loci. also, phage attachment sites frequently are located in this region. however, certain genes, and infrequently intergenic regions in operons are used by pai. in members of the enterobacteriaceae, the selC locus is an insertion site frequently used by functionally different pai in e. coli, shigella spp., and s. enterica (fig. 4). the overlapping sequences of tRNA loci and pai are within the 3’ end of the tRNA genes, are usually 15 to 20 nucleotides long and encode the 3’ side of the acceptor-TφC stem-loop region of a tRNA up to the conserved CCA end (156). the molecular basis of the use of tRNA genes as integration sites is not fully understood, but three hypotheses are plausible.

specific tRNAs are associated with a pai, so that the encoded tRNA may be used to read codons of the associated pai. this has been shown for the leuX tRNA gene encoding the rare tRNALeuX (292). expression of leuX is necessary for the synthesis of virulence factors encoded on paiLeuX. since basic cellular genes that are not involved in pathogenicity also are modulated by leuX and since the association of pai with specific tRNA genes is not found in other islands, this thesis is not favored.

a second hypothesis would include the presence of multiple copies of tRNA genes, providing multiple insertion sites and amplification of pathogenicity factors. this is, however, not true for selC and leuX, which occur in single copies.

the third, and most plausible, hypothesis suggests that the conserved structure in tRNA genes provides structural motifs that facilitate the integration and excision of pai and also
phages (290). This emphasizes that integration and excision are catalyzed by integrases.

Hou (156) proposed a fourth hypothesis, in which the 3 end of tRNA plays a major role. In his hybrid theory, the conserved CCA ends provide the initial site for integration by an integrase. The 3 end of a tRNA hybridizes to one strand of a duplex DNA during recombination. This stabilizes the separation of the DNA duplex for recombination (for details, see reference 156). Whether this theory or one of the others is correct has yet to be elucidated. Nevertheless, it is apparent that phages and PAI use conserved genes as integration sites. These conserved genes might confer safety to the mobile genetic element that they can integrate in any genome of members of a given population. This need could be, in an evolutionary biology point of view, important to maintain pathogenicity factors in a bacterial population (329).

PAI OF GRAM-NEGATIVE PATHOGENS

Helicobacter pylori

H. pylori infects the mucosa of the stomach, an organ that has long been considered an environment too hostile for bacterial colonization. Infections with H. pylori are common and are often acquired in childhood, and acute infection can lead to chronic colonization of the gastric mucosa (for a recent review, see reference 342). This colonization usually leads to chronic gastritis, and subsequent forms of disease are dependent on host as well as on bacterial factors. In the majority of individuals with gastritis, the infection remains asymptomatic. However, patients with low or high production of gastric acid can develop gastric ulcer or duodenal ulcer, respectively. There is also a strong correlation between infection with H. pylori and development of mucosa-associated lymphoid tissue lymphoma and gastric cancer, resulting in the classification of H. pylori as a carcinogen.

H. pylori organisms are curved, rod-shaped bacteria with a group of polar flagella and are covered by a membrane sheath. Motility is an important virulence factor and enables the bacteria to penetrate the mucin layer of the gastric epithelium (171). The bacteria also produce urease. This enzyme catalyzes the formation of CO2 and ammonia that can neutralize the acidic pH in the vicinity of the bacteria. Cultivation of H. pylori requires a microaerophilic atmosphere and complex media.

Clinical isolates of H. pylori have been classified into type I and type II strains, which are associated with different clinical outcomes ranging from gastric ulcer to asymptomatic colonization. There are also various forms of intermediate virulence. Type I strains carry genes encoding both, the cytotoxins CagA and VacA, while type II strains contain vacA genes only (376). VacA is a secreted toxin that induces extensive vacuolation in epithelial cells, cell death, and destruction of epithelial integrity.

The attachment of type I strains to gastric epithelial cells induces the synthesis and secretion of several chemokines, and the secretion of interleukin-8 (IL-8) is frequently assayed in model systems. It has also been observed that the infection of epithelial cells by H. pylori leads to dramatic rearrangements of the host cell actin cytoskeleton and the formation of pedestals (318) that are reminiscent of EPEC-induced pedestals, as well as to changes in the gross morphology of host cells (hummingbird phenotype). These phenotypes are associated with alterations in the signal transduction pathways of the host cell and the presence of a tyrosine-phosphorylated protein (317).
**cag PAI**. Detailed analysis of the cagA loci in type I and type II strains indicated that the latter group showed deletions of a large chromosomal region. This locus had the typical characteristics of a PAI and was termed the cag PAI (49). Censini et al. (49) characterized this locus and showed that the cag PAI had a size of 37 to 40 kb, flanked by direct repeats of 31 bp (Fig. 5A). The locus has a G+C content of 35%, in contrast to the 39% observed for the core genome (349). A gene for a tRNA has not been identified at the point of integration, but the glr gene (glutamate racemase) was disrupted by insertion of the PAI. There are no genes associated with DNA mobility within the cag PAI of type I strains. However, the presence of an IS 605 element within the cag PAI of strains with an intermediate virulence phenotype was observed. In strains of intermediate virulence, various forms of deletions with the cag PAI were detected, and in certain strains the locus was separated into two portions, referred to as cagI and cagII (1, 49). These observations support a correlation between the presence and integrity of the cag PAI and the severity of disease. Studies with a mouse model have shown that an association between cag PAI-negative *H. pylori* strains and cag PAI-positive strains that are mouse adapted and have modulated their ability to activate a proinflammatory response can better colonize mice than the parental strains do, indicating that the cag PAI of type I strains may become lost during colonization of infected animals (272). In addition to large deletions and chromosomal rearrangements of the cag PAI, there are indications that point mutations in the PAI genes result in ablation of CagA translocation and IL-8 induction (92). This effect can be explained by loss of function of the T4SS. Work by several groups has demonstrated that the translocation of CagA into target cells is required for these phenotypes (8, 11, 262, 316, 338). After translocation, CagA is tyrosine phosphorylated and induces growth factor-like phenotypes in the host cell. SHP-2 (SRC homology 2 domain [SH2]-containing tyrosine phosphatase) was identified as a cellular target of CagA (150, 319). It was observed that SHP-2 and CagA form a complex that could activate particular pathways and lead to actin polymerization and pedestal formation (150). Activation of SHP-2 by CagA might contribute to the abnormal proliferation and movement of gastric epithelial cells, thus contributing to the pathogenesis of *H. pylori* gastric infections. It has also been proposed that phosphorylated CagA may trigger the transcription of nuclear genes, which may explain the

FIG. 5. Examples of PAI of various pathogens. The topology of PAI of various pathogens is depicted to demonstrate different features of PAI. The functional classes of the genes are as indicated in the figure. (A) The cag island of *H. pylori* harbors genes for a type IV secretion system (T4SS) (grey symbols) that can mediate the translocation of the effector protein CagA (dark grey) into eukaryotic cells modified from reference 92. (B) *Salmonella* SPI-2 has a mosaic structure. It has been defined as a genetic element of about 40 kb that is absent from the related species *E. coli*. Only a 25-kb portion is required for systemic infection and encodes a T3SS system (grey), secreted proteins (dark grey), and regulatory proteins (white). Another portion (15 kb) is not required for virulence and harbors genes for metabolic or unknown functions (light grey symbols), such as an enzyme system for alternative electron acceptors during anaerobic growth. Genes associated with mobility are indicated by dark dotted symbols. Modified from reference 134. (C) The HPI of *Y. enterocolitica* is an example of an unstable PAI. Several elements are present within this PAI (dotted arrows). Genes in HPI encode an high-affinity iron uptake system (dark grey) that is important for the extracellular proliferation of the pathogen during colonization of the host. Modified from reference 45. (D) The *vSal* PAI of MRSA is shown. A remarkable feature of PAI in *S. aureus* is the presence of a large number of genes with related functions, such as genes for enterotoxin (dark grey) or lipoproteins (grey). Modified from reference 9.
increased frequency of gastric cancer in patients infected with *cagA*-positive *Helicobacter* strains (75).

Sequence analysis of the *cag* PAI indicated that a T4SS is encoded by this locus and that CagA is a translocated substrate of the secretion apparatus. Only 6 of the 27 to 29 predicted open reading frames (ORF) in the *cag* PAI show significant sequence similarity to components of the T4SS of other bacteria, and the contribution of other genes to formation of a translocation apparatus is not clear (Fig. 5A).

Systematic mutagenesis approaches to analysis of the functions of the genes in the *cag* PAI were performed by Fischer et al. (92) and Selbach et al. (320). The individual inactivation of 27 putative genes and phenotypic analysis identified a subset of 17 genes that are required for the translocation of CagA into host cells and a subset of 14 genes that are required for the stimulation of IL-8 synthesis in host cells. Although the assembly of T4SS is not understood in full detail, these observations indicate that the majority of genes within the *cag* PAI are required for the formation of a functional T4SS, by encoding either structural components or protein important for the assembly and regulation of the system. Neither approach resulted in the identification of mutant strains that were deficient in CagA translocation but capable of inducing IL-8 secretion. These observations indicate the absence of a further translocated protein responsible for IL-8 induction within the *cag* PAI or a direct effect of the T4SS in IL-8 induction. The secretion of VacA is not dependent on the *cag* PAI, and so far no further proteins translocated by *cag* PAI-encoded T4SS have been identified.

The observation that the *cag* PAI is absent or partially deleted in *H. pylori* strains with low virulence might suggest that the function of the *cag* PAI-encoded T4SS is not compatible with a long-lasting colonization of the gastric epithelium. The inflammatory response elicited by *H. pylori* after contact-dependent translocation could lead either to a clearance of the infection or to a severe immunopathology. However, epidemiological data indicate that the frequency of *cag* − isolates of *H. pylori* is much higher in the Asian population than in the Western population, indicating that further host and pathogen factors are involved in colonization (21).

The genome of *H. pylori* is characterized by a high flexibility, and an extremely high frequency of recombination was observed (89, 341). The DR flanking the *cag* PAI probably function as sites for recombination and deletion of the locus.

**Pseudomonas aeruginosa**

*Pseudomonas* spp. are widely distributed in nature and occur in both soil and aquatic habitats. They have a large metabolic versatility and are able to utilize numerous substrates as carbon and energy sources. *Pseudomonas aeruginosa* is well known as an opportunistic pathogen for plants, animals, and humans (42, 297). A variety of human infections ranging from superinfections of mucosal tissues such as the urinary and respiratory tracts. A predisposing condition for manifestation of *P. aeruginosa* infections is a breach in the host immune system or the specialized nature of the underlying disease, such as cystic fibrosis. *P. aeruginosa* is involved in a significant number of cases of urinary infections in patients with indwelling catheters and is a nosocomial pathogen. Many *P. aeruginosa* infections are difficult to treat since this organism can express multiple antibiotic resistance factors (28, 53, 97, 114, 118, 358).

The 6,264,403-bp chromosome of *P. aeruginosa* strain PAO1 has been sequenced completely (340). Besides a large gene repertoire involved in the catabolism, transport, and efflux of organic compounds, which basically is responsible for its metabolic versatility, *P. aeruginosa* possesses 10 islands of 3.0 kb and larger, which have a significantly lower G+C content than the rest of the chromosome (66.6% for the rest of the chromosome) and show an unusual codon usage (203).

Some islands carry apparently dispensable genes that are not present in all *P. aeruginosa* strains, such as genes encoding toxins, pyocins, and proteins with unknown functions. Other islands encode cellular appendages and elements of the outer membrane such as lipopolysaccharide (LPS) biosynthesis enzymes. These islands are referred to as PAI-like structures, since analysis to determine the role of these elements have not been performed so far.

*P. aeruginosa* displays interclonal heterogeneity. Comparison of the genomes of *P. aeruginosa* strain PAO1 and *P. aeruginosa* strain C has shown that the latter possesses 700 kb of additional DNA and carries 11 regions of 5 kb and larger (20 to 160 kb) which are not present on PAO1. These regions have a lower G+C content and may be considered to be PAI.

In *P. aeruginosa* strain PAO1, a T3SS is encoded by the 25,670-bp exoenzyme S regulon, with subunits displaying a high level of sequence similarity to the components of the *Yersinia* Yop virulon. The T3SS permits contact-mediated translocation of the antihost factors ExoS, ExoT, ExoU, and PcrV. ExoS and ExoT are related proteins that have 75% amino acid identity (101). The ExoS protein uncouples the Ras-mediated signal transduction pathway. While the C terminus of ExoS possesses ADP-ribosyltransferase activity, the N-terminal domain is responsible for the disruption of actin and is therefore cytotoxic (101, 269). ExoT also possesses ADP-ribosyltransferase activity, but only 0.2% of the amount in ExoS. Recently, it has been shown that ExoT functions as anti-internalization factor which prevents uptake by multiple cell lines and is able to modify the host cytoskeleton (112). Expression of another antihost factor, ExoU, correlates with acute cytotoxicity and lung injury. ExoU was also demonstrated to cause necrosis of macrophages. Moreover, the T3SS transports at least another macrophage-killing activity, which is independent of ExoU and causes apoptosis (139). PcrV is only poorly characterized. It is thought to be involved in modulation of the host cytokine response. Apparently, the T3SS also causes ExoU-independent oncisis of macrophages and polymorphonuclear leukocytes, cellular and nuclear swelling, disintegration of the plasma membrane, and absence of DNA fragmentation. The ExoS regulon may be considered an ancient PAI that became irreversibly fixed in the genome and has lost all elements of mobility.

**PAGI-1.** Differential hybridization of a *P. aeruginosa* isolate from a patient with urinary tract infection and strain PAO1 resulted in the discovery of *P. aeruginosa* genomic island 1 (PAGI-1). It consists of 48,893 bp of DNA and contains 51
ORF. The G+C content of PAGI-1 has an asymmetric distribution. In the first three-quarters of the sequence, from ORF 1 to ORF 30, the G+C content is 63.7%, and the remaining portion of the PAI, from ORF 31 to ORF 51, has a G+C content of 54.7%, which is lower than that of the P. aeruginosa core chromosome. Approximately 50% of ORF located on PAGI-1 encode hypothetical proteins with unknown functions. Among the genes that could be assigned to putative functions, the most notable are remnants of transposable elements, putative transcriptional regulators, and a number of genes encoding various dehydrogenases. At the left end of PAGI-1 are two ORF encoding proteins homologous to paraquat-inducible proteins of E. coli. In E. coli, a pair of paraquat-inducible genes, pqaA and pqaB, are under the control of the SoxS and SoxR regulators, which respond to redox-cycling agents capable of generating superoxide radicals in the cell. The precise role of PqaA and PqaB in detoxification of radicals is not known. The presence of these genes in P. aeruginosa and the large number of dehydrogenases with putative functions in the detoxification indicates a role for this island in the detoxification of reactive oxygen species, i.e., in protection against oxidative damage (209). Biological studies to confirm this suggestion are under way.

PAGI-2 and PAGI-3. Further analyses of P. aeruginosa strains C (isolate from the airways of a patient with cystic fibrosis), SG17M (an isolate from the aquatic environment), and PAO1 (an isolate from a patient with cystic fibrosis) revealed two strain-specific islands that are integrated into tRNA^{Gly} genes, which are located within a cluster comprising one tRNA^{Gly} gene followed by two identical tRNA^{Gly} genes. The first island is designated PAGI-2(C), is present in P. aeruginosa C, and consists of 104,996 bp of DNA; the second is designated PAGI-3(SG), is present in SG17M, and consists of 103,304 bp of DNA. Analysis of PAGI-2(C) and PAGI-3(SG) sequences revealed 111 and 106 ORF, respectively (204). In both strains, the gene islands are partitioned into two larger blocks. The cluster adjacent to the attL site harbors strain-specific genes. The other gene block predominantly contains hypothetical genes with various homologues, e.g., to genes in Xylella fastidiosa islands. Some of the strain-specific genes of PAGI-2(C) putatively encode proteins for complexation and transport of heavy metal ions, essential proteins for cytochrome c biogenesis and related thiol disulfide exchange proteins. Moreover, proteins for cation transport, a two-component regulatory system, transcriptional regulators, a transport system conferring mercury resistance, and other transporters are present. The role of these proteins is unclear. The authors hypothesized that cytochrome biogenesis may facilitate iron uptake and inactivation of peroxides and thus may confer advantage to the bacteria in persistence in the lungs of cystic fibrosis patients, where they suffer from iron limitation and oxidative stress (204). Other genes, such as those responsible for copper homeostasis and mercury resistance, may be important not for pathogenesis but for survival in an environment with high heavy-metal concentrations. PAGI-3(SG) of the environmental isolate SG17M is thought to be a metabolic island containing genes related to metabolism and transport of amino acids, coenzymes, porphyrins, and a number of putative enzymes (204).

Besides PAGI-1, PAGI-2(C), and PAGI-3(SG), large (100-kb) islands that are derived from plasmids and reversibly recombine with either of the two tRNA^{Gly} genes have been found in isolates of P. aeruginosa (184).

Glycosylation island. Recently, a glycosylation island was discovered in P. aeruginosa strain PAK (7). This island is obviously present in all strains that express a-type flagellin and mediates posttranslational glycosylation of this type of flagellin by covalent attachment of glycosyl moieties to one or several sites within the polypeptide. This island is ~16 kb in size and contains 14 ORF; most of these may be involved in glycosylation and various biosynthetic pathways, and some are of unknown function. The precise role of glycosylation in flagellar function is not known. Glycosylation is not necessary for motility under laboratory conditions, but the strong association of this island with certain pathogenic strains of P. aeruginosa suggests that there are some functions for glycosylation in vivo.

*Shigella* spp. *Shigella* spp. are gram-negative, facultatively anaerobic rods that can be divided into the four species *S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*. *Shigella* spp. are the causative agents of shigellosis (bacillary dysentery) (273). Infection with *Shigella* spp. causes a spectrum of clinical outcomes ranging from mild watery diarrhea to classic dysentery with fever, intestinal cramps, and discharge of mucopurulent and bloody stools. In addition, inflammation of the infected tissue is a hallmark of shigellosis (224).

*Shigella* spp. are able to invade epithelial cells via M-cells to gain entry into the colonic epithelium. After transcytosis, the bacteria enter lymphoid follicles containing resident tissue macrophages. The cells are engulfed by macrophages but not destroyed. After liberation from the phagosome into the host cell cytoplasm, they cause caspase-1-mediated apoptosis and release of IL-1β and IL-8. The inflammatory response to these cytokines damages the colonic mucosa and exacerbates the infection. After apoptotic release of bacteria from infected macrophages, *Shigella* gains access to adjacent enterocytes via the basolateral side. This pathogen cannot enter polarized epithelial cells from the apical side.

The genes for invasiveness are located on the large invasion plasmid (225, 299). This plasmid contains a 30-kb DNA region encoding a T3SS including Mxi (for “membrane excretion of Ipa”) and Spa (for “surface presentation of invasion plasmid antigen”) proteins, and a second operon encoding effector proteins such as Ipa (for “invasion plasmid antigens”). In addition to this plasmid, chromosomal genes are needed for the full array of virulence phenotypes caused by *Shigella* spp. So far, five PAI have been identified in *Shigella* spp. Although these PAI carry virulence genes, their contribution to pathogenesis is not yet fully understood (164, 217, 279, 282, 361).

**SII-O. LPS** is an important virulence factor of *Shigella* spp. (210). Since the immune response to *Shigella* spp. is O-antigen specific, an immune response to a specific O antigen does not protect against infection with other serotypes. Therefore, the capacity to alter serotypes may be of advantage for *Shigella* spp. in the infectious process.

More than 13 *S. flexneri* serotypes are known, basically defined by differences in the O-antigen structure. The difference
in serotype is caused by glucosylation and O acetylation of the basic O antigen. These modifications are catalyzed by enzymes, which are encoded predominantly by bacteriophages. The only known exceptions are the genes for determination of O antigen in serotype 1, which are obviously located on a PAI termed SHI-O (164).

The gene composition and order indicate that this PAI is derived from an ancestor bacteriophage which has lost its ability to be excised from the bacterial chromosome. A larger part of the phage genome is deleted. The remaining genome is flanked by two typical phage attachment sites which are located in an unusually short distance of 6.5 kb. Interestingly, the G+C content of SHI-O is around 40%, much lower than the G+C content of the rest of the Shigella chromosome (49 to 53%). Besides phage gene remnants and insertion elements, the SHI-O island contains three ORF whose products have high sequence identity to proteins encoded by other serotype-converting bacteriophages. The respective gene products are putatively involved in glucosylation reactions necessary for serotype conversion. The first gene, gtrA, encodes a highly conserved, hydrophobic 120-amino-acid integral membrane protein with unknown function. It may function as flippase for the UndP-glucose precursor (124). The gtrB gene encodes a bacteroprenol glucosyltransferase which may catalyze the transfer of glucose from UDP-glucose to bacteroprenol phosphate, and the third gene, gtr, determines a serotype-specific glucosyltransferase (5).

**SHI-1.** The second PAI, first described in *S. flexneri* serotype 2a, has been completely sequenced from strain YSH6000T. This PAI was originally termed “she” because it contains the she gene (282). It consists of 46,603 bp, is located directly downstream of the pheV rRNA gene, and includes an imperfect repeat of the 3′-end 22 bp of the pheV gene at the right boundary of that PAI. SHI-1 contains a bacteriophage P4-like integrase gene, intact and truncated mobile genetic elements, plasmid-related sequences, ORF similar to those found in the EHEC LEE and SHI-2 and in the sigA, pic (she), set1A, and set1B genes, as well as two novel ORF (4, 282). Several of the proteins encoded by SHI-1 are thought to be virulence factors. SigA and Pic (also known as ShMU) belong to the autotransporter family of proteins (143). The Shet1 protein is an enterotoxin of the CT/LT-like toxin family. SigA (*Shigella* IgA-like protease homologue) is a 139.6-kDa protein exhibiting protease activity and cytopathic effects on HEP2-cells. Mutants with mutations in sigA caused 30% less fluid accumulation in a rabbit loop model of infection, suggesting that this protein plays a role in *Shigella* pathogenesis.

The Pic (for “protein involved in intestinal colonization”) protein is a serine protease of 109.8 kDa, which is able to cleave gelatin as well as bovine and murine mucin. The mucus layer overlying the mucosal surface is considered to be a protective barrier against enteric infections (283). Some enteric pathogens have developed strategies to penetrate this layer, such as flagella motion or production of mucus-degrading enzymes (81, 103). It has been speculated that Pic could be involved in enabling *Shigella* spp. to burrow through the mucin layer, helping to establish initial colonization (141). Other functions of Pic contribute to serum resistance and hemagglutination activity.

The set1A and set1B genes encode the enterotoxin Shet1, which also increases fluid accumulation in the rabbit loop model. It is suggested that Shet1 is involved in the development of watery diarrhea. A novel ORF designated sap (for “*Shigella* autotransporter-like protein”) has 87% identity to AG43, an autotransported surface protein of *E. coli* that mediates autoaggregation. However, the biological role of Sap in *Shigella* is not clear (142).

**SHI-2.** The third PAI, SHI-2, was discovered independently by two groups (244, 361). It is 23.8 kb long and is inserted at the setC RNA locus in *S. flexneri*. It carries a couple of cryptic genes, an aerobactin operon, a colicin V immunity gene, and several novel ORF without known function. Only for the aerobactin operon and the colicin V immunity gene could an association with virulence be made. The aerobactin operon found in SHI-2 is similar to that found in the *E. coli* pColV-K30. In general, SHI-2 appears to facilitate the survival of *S. flexneri* in stressful environments. The aerobactin contributes to the survival and growth of SHI-2-harboring bacteria in low-iron environments (60, 65), and immunity to colicin V allows these strains to survive in environments where they compete with other bacteria. Two other putative virulence factors were proposed to be encoded on SHI-2: a tetracycline transporter and a protein involved in scavenging free oxygen radicals.

**SRL.** The *Shigella* resistance locus (SRL) was discovered following the spontaneous loss of multiple antibiotic resistance by *S. flexneri* 2a strain YSH6000. This loss was accompanied by the deletion of a chromosomal region of about 99 kb. Within this element, a smaller (66-kb) element carries the multiple antibiotic resistance determinants (217).

SRL is 66,257 bp in length and harbors 59 ORF. It is integrated downstream of the serX rRNA gene, contains an integrase gene adjacent to serX, and further genes encoding resistance to streptomycin, ampicillin, and tetracycline and a complete ferric dicitrate uptake system. Interestingly, a couple of genes are related to the CP4 prophage group and the 933L prophage. Besides enterobactin and aerobactin, this is the third siderophore system described in *Shigella* spp. and the second system encoded on a PAI (217).

**SHI-3.** The SHI-3 island of *S. boydii* maps between pheU and yidA and is 21 kb in size. It contains a P4-like integrase gene, an ORF with 97% sequence identity to shiB, an ORF with similarity to the ORF found in SHI-2, a functional aerobactin operon, IS elements, and prophage genes homologous to those found in the LEE of *E. coli* O157:H7. Although it carries an aerobactin operon with high sequence identity to that found in SHI-2, SHI-3 is distinct from SHI-2 (279). There is a deletion of more than 6 kb relative to the *E. coli* K-12 sequence at the site occupied by SHI-3, including the lysine decarboxylase gene cadA. Cadaverine, produced by the decarboxylation of lysine, inhibits *Shigella* enterotoxic activity, and deletion of cadA was demonstrated to enhance the virulence of *Shigella* (226). Therefore, acquisition of SHI-3 may be associated with virulence in *S. boydii*.

**Other Shigella islands.** Sequencing of the *S. flexneri* 2a genome has given us some new insights into the genome structure and the presence of further putative PAI (168). The chromosome shares a backbone sequence of ~3.9 Mb with *E. coli* K-12 (MG1655) and O157 (EDL933); the gene order is essentially colinear. This backbone is, however, interrupted by nu-
merous segments of *E. coli* K12-, *E. coli* O157-, and Shigella-specific segments designated K-islands, O-islands, and S-islands, respectively. The sequenced strain SF301 contains 64 S-islands larger than 1 kb, among which SHI-1 and SHI-2 are present. In addition, seven *ipaH* genes were found in five large *Shigella*-islands, which are thought to be originally phage transmitted. These islands were designated *ipaH* islands. *ipaH* island 2 may be associated with iron uptake, and two other islands, the *sci* and SII islands, are worth mentioning since they have similarities to the *Salmonella* sciCDEFF operon and are involved in the expression of a type I antigen, respectively. The functions of these islands have not been investigated in detail (168).

**Yersinia** spp.

The genus *Yersinia* includes three species that are human pathogens. *Y. pestis*, the causative agent of plague, has caused several epidemics with a large number of fatal cases. While plague is no longer a major health problem with respect to naturally occurring infection, today *Y. pestis* is considered a potential agent of bioterrorism. In contrast, the species *Y. enterocolitica* and *Y. pseudotuberculosis* still cause a significant number of gastrointestinal infections as a result of ingestion of contaminated meat products. Such infections are often associated with secondary episodes such as reactive arthritis. There are highly virulent serotypes of *Y. enterocolitica* and *Y. pseudotuberculosis*, as well as serotypes with low virulence, which are found as environmental bacteria. An import virulence factor of pathogenic *Yersinia* species is the virulence plasmid pYV, which encodes a T3SS and various translocated effector proteins and has been designated an “antihost genome” (58).

**HPI.** The full virulence characteristics are expressed only by *Yersinia* spp. that harbor the high-pathogenicity island (HPI) (for a review, see reference 45). HPI is present in *Y. pestis* and highly virulent serotypes of *Y. pseudotuberculosis* and *Y. enterocolitica* but absent in low-virulence serotypes. Analyses of HPI in the three pathogenic species indicated that the locus is between 36 and 43 kb in size and has a G+C content of 60%, which is higher than the average G+C content of 46 to 50% for the core genome of *Yersinia* spp. The HPI of all three species are flanked by a gene for tRNA^Asn^; however, the other flanks of HPI is less well defined and varies among the species. The functions of genes within a conserved portion of HPI of about 30 kb have been characterized and indicate that an iron uptake system is encoded by HPI (Fig. 5C). During the colonization of an eukaryotic host, iron is one of the most limiting factors for bacterial proliferation, and a variety of systems are found in bacteria for the synthesis of high-affinity iron-binding molecules (siderophores) and their binding and uptake. The HPI-encoded system appears to be especially important for the extracellular lifestyle of *Yersinia* spp. within the infected host. Enzymes encoded by HPI synthesize the siderophore yersiniabactin, and further genes within this PAI encode a membrane-bound transport system of iron-loaded yersiniabactin (286). A further gene encodes a regulator for the expression of structural genes. In addition to the strong correlation between the presence of HPI and the relative virulence of various *Yersinia* spp., experimental analyses indicated that the inactivation of the HPI-encoded outer membrane yersiniabactin receptor FyuA (286), or the *irp* genes for biosynthesis of the siderophile in *Y. enterocolitica* (46) as well as *Y. pestis* (16), resulted in dramatically attenuated virulence of the mutant strains in a murine infection model.

HPI is a genetically unstable PAI with a frequency of loss of $2 \times 10^{-3}$ in *Y. pestis* and $1 \times 10^{-4}$ in *Y. pseudotuberculosis*, but it appears to be more stable in *Y. enterocolitica* (45). Several DRs, copies of the IS700 element (in *Y. pestis* and *Y. pseudotuberculosis*) or IS1328, IS1329, and IS400 (in *Y. enterocolitica*), and a functional integrase gene have been identified in unstable versions of HPI. Lack of the integrase gene and a DR could explain the relative stability of HPI in *Y. enterocolitica*. In all three species, the HPI is inserted at genes for tRNA^Asn^, However, HPI of *Y. pestis* and *Y. pseudotuberculosis* have integrated at different genes for tRNA^Asn^, indicating independent acquisition of the locus by distinct evolutionary lineages of the genus *Yersinia* (38, 137). Indeed, the integration of a cloned segment of the *Y. pestis* HPI has been experimentally demonstrated and was not selective for a specific tRNA^Asn^ gene (284).

The genetic elements that contribute to the instability of the locus may also contribute to the mobility of HPI and its distribution to other members of the *Enterobacteriaceae*. Recently, it has been observed that the HPI is also present in various other enterobacterial pathogens, indicating that the HPI-encoded iron uptake system is important for the pathogenic potential of many bacteria (313). This group of pathogens include septicemic, uropathogenic, and diarrheagenic strains of *E. coli* (such as EPEC, EAEC, EIEC, ETEC, and STEC), as well as diffuse-adhering *E. coli* (DAEC), *Citrobacter* spp., *Klebsiella* spp., *Enterobacter* spp. (63, 175, 313). HPI was not detected in *Salmonella enterica* subspecies I, which comprises the majority of human isolates, but is present in *S. enterica* subspecies III and IV (264). In *Yersinia* spp., the HPI is inserted at the *asnT* tRNA gene, and the same insertion point was identified in most of the other species harboring HPI. The HPI in *Yersinia* spp. and other pathogens are highly conserved; indicating that distribution of this locus is a rather recent evolutionary event.

**Chromosomally encoded type III secretion systems.** Further PAI-like chromosomal regions have been identified recently in *Yersinia* spp., but their role in virulence is not fully understood. In addition to the T3SS encoded by the virulence plasmids pYV of *Yersinia* spp., there are chromosomal loci that encode additional T3SS in all three highly virulent species of *Yersinia*. In contrast to pYV-encoded T3SS genes, which are closely related among the three species, the chromosomal T3SS loci appear to have a different genetic origin (98). Genome sequencing of *Y. pestis* also revealed the presence of genes encoding a second T3SS (267). Interestingly, this locus is not related to the ysa genes of *Y. enterocolitica*, since the corresponding gene products are more similar to SPI-2 genes of *S. enterica*. Further work is needed to clarify whether the chromosomal T3SS locus of *Y. pestis* contributes to specific virulence characteristics of *Y. pestis*, for example during the infection of rodents or fleas as the vector. It is possible that the life-style of *Y. pestis* in these hosts is different from that in human hosts, thus requiring additional virulence functions.
**Vibrio cholerae**

*V. cholerae* is a comma-shaped, highly motile species that can be found as a normal inhabitant of aquatic habitats. However, it can also convert to a highly pathogenic human pathogen. Infections with *V. cholerae* lead to cholera, a disease with a history of pandemics and a large number of casualties. Today, cholera is still a major health risk in developing countries and in communities with low standards of hygiene.

The interaction between *V. cholerae* and the host is restricted to the intestinal mucosa, and the pathogen does not invade the epithelial layer or deeper tissue. Cholera is characterized by diarrhea and vomiting, resulting in massive loss of fluid and electrolytes that can rapidly lead to death. The cholera toxin is the major virulence determinant of *V. cholerae* and is responsible for the diarrheal symptoms. Infection of non-toxigenic strains by filamentous phage CTXΦ harboring genes encoding cholera toxin has resulted in the appearance of toxigenic strains. This lysogenic conversion led to the emergence of new pathogenic strains (364).

**VPI.** The presence of a receptor for CTXΦ phage binding is required for lysogenic conversion, a process that has resulted in the appearance of new toxigenic strains of *V. cholerae*. Molecular genetic analyses identified VPI (for “Vibrio cholerae pathogenicity island”) (173) with an essential function in this process. VPI is about 40 kb long and has a G+C content of only 35% compared to an average G+C content of the *V. cholerae* genome of 47 to 49%. Furthermore, VPI is inserted adjacent to ssrA, a gene with similarity to rRNA genes, and is flanked by *att* sites. Only one of these *att* sites remains after loss of VPI by naturally occurring deletions of the locus. VPI encodes the toxin-coregulated pilus (TCP), a type IV pilus acting as an essential colonization factor by contributing to the adhesion of *V. cholerae* to the epithelial surface. In addition, TCP functions as the receptor for CTXΦ, the bacteriophage harboring the cholera toxin genes (364). Thus, the presence of VPI is a prerequisite for the lysogenic conversion of nontoxigenic *V. cholerae* strains by CTXΦ, an event that has contributed to the appearance of new toxigenic strains that caused cholera pandemics.

The identification of a toxigenic *V. cholerae* strain without VPI initiated the investigation of the mobility of VPI. Detailed analysis of VPI indicated the presence of a large number of bacteriophage genes, and Karaoğlìs et al. (174) reported that VPI is the chromosomally integrated form of a bacteriophage termed VPIφ. These authors also demonstrated that several VPI-positive strains produce intact phage particles of VPIφ that can infect VPI-negative strains. A structural component of TCP, the TcpA subunit, has sequence similarity to coat proteins of bacteriophages. In fact, binding of antibodies against TcpA to isolated VPIφ phages was observed, indicating that TcpA has a dual function as pilus subunit and structural phage protein. VPI also encodes ToxT, a transcriptional activator within the regulatory cascade controlling the expression of cholera toxin, TCP, and further genes. A further virulence factor encoded by VPI is Mop, a putative protease modulating the virulence of *V. cholerae*. A strain defective in *mop* was hypersensitive in the rabbit ileal loop model and caused more severe mucosal symptoms than did wild-type *V. cholerae* (381).

The VPI of *V. cholerae* is also an excellent example of the contribution of bacteriophages to the distribution of PAI. This PAI is identical to the genome of bacteriophage VPIφ, and chromosomally integrated VPI still can produce infective phage particles. VPI also demonstrates that certain bacteriophage genes have functions in the life cycle of the phage itself, such as forming the phage coat, and functions that are beneficial to the bacterial host, such as a function in adhesion. Bacteriophages conferred functions to their hosts that increased their ability to colonize host organisms, an observation that extends the coevolution of bacteria and bacteriophages beyond a predator-prey relationship.

The integration of bacteriophages in the genome of *V. cholerae* is most probably a very recent event, and it can be assumed that horizontal transfer of other PAI occurred by similar mechanisms. The analysis of *V. cholerae* isolates from the seventh pandemic and prepandemic isolates indicated that the acquisition of VPI was an important step for the emergence of a *V. cholerae* strains with extended virulence potential (78).

**VPI-2.** Recently, a second PAI in *V. cholerae* was identified and was termed VPI-2 (167). VPI-2 was detected in toxigenic but not in nontoxigenic strains of *V. cholerae* serogroup O1. This PAI of 57.3 kb is inserted at the gene for tRNAAsp and contains DR and a bacteriophage integrase. The accessory virulence factor neuraminidase is encoded by VPI-2. Neuraminidase is secreted and converts oligosaccharides by cleavage of sialic acid. This reaction results in a larger number of ganglioside Gm1-like molecules on host cell surfaces, which serve as binding sites for the cholera toxin. However, this process is not absolutely required for the action of cholera toxin (110).

**VSP-I and VSP-II.** Outbreaks of cholera can show epidemic as well as endemic forms, and recent microarray analyses have shown that strains of the seventh cholera pandemic have acquired genes that are not present in pre-seventh-pandemic strains. The appearance is considered a prerequisite for spread of strains that caused the seventh pandemic, starting in 1961. Hybridization analysis of genomic arrays based on the genome sequence of a *V. cholera* serotype O1 strain from the seventh pandemic (78) identified two genomic regions in the pandemic strain that were absent in prepandemic strains; they have been termed “Vibrio seventh-pandemic island I” (VSP-I) and VSP-II. These loci show several characteristics of PAI, including a low G+C content. The G+C content of VSP-I is 40%, in contrast to 47% for the entire genome. Beside containing a number of genes without known functions, this island putatively encodes a deoxycytidylate deaminase-related protein, a transcriptional regulator, and a protein with phospholipase A activity (78). VSP-II comprises eight ORF with hypothetical and conserved functions, which are not well defined. A role for VSP-I and VSP-II in virulence has not been reported so far.

**Salmonella spp.**

*S. enterica* is one of the most common food-borne bacterial pathogens. Disease outcome ranges from localized self-limiting gastrointestinal infections to typhoid fever, a life-threatening systemic bacteremia. Classic taxonomy has classified *Salmonella* isolates into over 2,000 species, but it is now agreed that these isolates are serotypes of two species with several
subspecies. Almost all clinical isolates are serovars of *Salmonella enterica* subspecies I, such as *S. enterica* serovar Typhimurium, Enteritidis, or Typhi. *Salmonella bongori* is the second, phylogenetically older species, which is only rarely associated with human infections. There is also a remarkable variation in host range of various serotypes of *S. enterica*, from highly host-specific serotypes to serotypes that infect a broad spectrum of hosts. Although many virulence factors have been identified in *S. enterica* serovars Typhimurium and Typhi, the underlying reasons for the different host specificities and disease outcome of various serovars are not fully understood. However, a role of horizontally acquired DNA fragments such as PAI has to be assumed (280).

Most virulence factors of *S. enterica* are determined by chromosomal genes, and many of these are located within PAI. The PAI of *S. enterica* serotype Typhimurium are referred to as *Salmonella* pathogenicity islands (SPI) (323). Work by the group of C. A. Lee first revealed that genes required for the invasion phenotype are clustered with a defined region of the chromosome of *S. enterica* (237), now termed SPI-1. SPI-1 and SPI-2 both encode T3SS for the translocation of virulence proteins into eukaryotic target cells, but the roles of SPI-1 and SPI-2 in virulence are entirely different.

**SPI-1.** The function of SPI-1 is required for the invasion of nonphagocytic cells, an important virulence trait of *S. enterica* (reviewed in reference 109). The SPI-1-encoded T3SS form needle-like surface appendages (197) that can mediate the delivery of proteins by extracellular *Salmonella* organisms into host cells. A set of effector proteins encoded by SPI-1 and additional genes outside of SPI-1 are translocated. One subset of these effector proteins modify signal transduction pathways resulting in the temporal reorganization of the actin cytoskeleton of the host cell. The structures and functions of SpIP, SopE, and SopE2 have been analyzed in detail. These proteins interfere with the function of host cell proteins of a family of small GTPases (e.g., Cdc42, Rac-1, and Rho) that regulate the formation of F-actin filaments and the dynamics of the cytoskeleton. In detail, SopE and SopE2 function as GTP exchange factors, resulting in activation of Cdc42 and leading to the formation of actin filaments at the site of translocation of these effector proteins (136). The localized modification of the cytoskeleton is followed by dramatic changes in the host cell surface that appear as membrane ruffles. As a consequence, nonphagocytic cells, in this case epithelial cells, internalize larger particles such as bacteria in a process termed macroinocytosis. It has also been demonstrated that the effector protein SpIP has an antagonistic effect by its function as a GTPase-activating factor. SpIP can mediate the inactivation of Cdc42 and Rac-1, resulting in the termination of actin polymerization and membrane ruffling (106). Translocation of SPI-1 effector proteins into macrophages induces a very rapid form of apoptosis. It was demonstrated that the SPI-1-encoded protein SipB is involved in macrophage apoptosis by activation of caspase-1 (149), a function similar to that of IpaB of *Shigella* spp. (149). This also results in the release of proinflammatory cytokines like IL-8. Caspase-1 induces apoptosis probably by activating other caspases (reviewed in reference 189).

Invasiveness appears to be one of the phenotypes related to SPI-1, and systemic infections can also be observed after oral ingestion of mutant strains deficient in the function of the SPI-1-encoded T3SS. It has been demonstrated that the function of SPI-1 is also related to the diarrheal symptoms (reviewed in reference 365). The second subset of effector proteins (SopA, SopB, and SopD) translocated by the SPI-1 system is required for this phenotype. SopB is an inositol phosphate phosphatase, and its enzymatic activity results in activation of chloride channels in the membrane of epithelial target cells, finally leading to the secretion of chloride and loss of fluid into the intestinal lumen (256). The function of SopA and SopD is not fully understood, but both effectors also contribute to the diarrheal phenotype in the bovine model of *Salmonella enteritidis* (169, 382).

SPI-1 is about 40 kb long and forms a *Salmonella*-specific insertion between genes that are consecutive in *E. coli* K-12 (237). This PAI is not associated with a tRNA gene, and the base composition of SPI-1 is 47% G+C, lower than the average G+C content of the core genome (which is about 52%). Genetic elements involved in DNA mobility are not obvious in SPI-1, and the locus appears to be stable in all clinical isolates of *Salmonella* spp. analyzed so far. However, the deletion of major parts of SPI-1 has been reported for environmental isolates of *S. enterica* serovars Senftenberg and Litchfield (115). Not all SPI-1 genes are related to the T3SS functions, since the *sit* gene cluster encoding an iron uptake system is also located within SPI-1 (386). The inactivation of this system had no effect on virulence, which is in line with the frequently observed redundancy of iron uptake systems. The regulation of SPI-1 expression is a complex process, which is not fully understood. SPI-1 expression responds to environmental stimuli and involves the SPI-1 encoded regulators HiiC, HiiD, HiiA, and InvF (details are given in Fig. 3A).

**SPI-2.** The function of SPI-2 is essential for the second hallmark of *Salmonella* pathogenesis, the ability to cause systemic infections and the proliferation within host organs. This virulence phenotype is linked to the ability of *S. enterica* to survive in phagocytic cells and to replicate within the *Salmonella*-containing vesicle in a variety of eukaryotic cells. SPI-2 was identified by selection of mutant strains that are unable to proliferate systemically in infected mice (148), as well as by screening for *Salmonella*-specific regions of the genome (261). Mutant strains deficient in the SPI-2-encoded T3SS are highly attenuated in virulence (323), and the use of such strains as vaccines against typhoid fever (152) or as carrier strains for recombinant vaccines (231) has been evaluated.

There are several cellular phenotypes related to SPI-2. The function of the SPI-2-encoded T3SS is required for the protection of the pathogen within the *Salmonella*-containing vesicle (SCV) against effector functions of the innate immunity. It has been reported that SPI-2 function prevents the colocalization of the phagocyte oxidase (357) as well as the inducible nitric oxide synthase (50) with the SCV. Both functions might be related to the modification of host cell trafficking by SPI-2 (353). As a consequence, intracellular *Salmonella* organisms are protected against damage by reactive nitrogen intermediates and reactive oxygen intermediates and against the potent antimicrobial activity of peroxynitrite, which is generated by the reaction of reactive nitrogen and oxygen intermediates. These defense mechanisms represent a specific adaptation to the intracellular environment, especially with phagocytic cells.
SifA, an effector protein encoded on a locus outside of SPI-2, is translocated by the SPI-2-encoded T3SS, and this protein is required to maintain the integrity of the phagosomal membrane of the SCV during intracellular proliferation. Several proteins encoded by SPI-2 genes are secreted (reviewed in reference 145) and translocated into host cells (198). For the SPI-2-encoded SpiC, translocation and a function in interference with cellular trafficking was reported (353); however, there is also evidence that SpiC is a functional component of the T3SS itself (102, 379). The SPI-2 locus is 40 kb long and is associated with the valV tRNA gene. At the same location, an insertion of 9 kb without virulence functions was detection in E. coli K-12 (147). The SPI-2 locus is 40 kb long and is associated with the valV tRNA gene. At the same location, an insertion of 9 kb without virulence functions was detection in E. coli K-12 (147). The SPI-2 locus is stable and appears to be conserved throughout various serovars of S. enterica. SPI-2 has a mosaic structure of at least two genetic elements. A portion of 25 kb encodes the T3SS and has a rather low G+C content of 43%. A further portion of 15 kb shows a base composition similar to that of the core genome, and genes in this portion are not required for T3SS function. This portion harbors genes encoding metabolic functions such as the tetrahydronate reductase (146) (see Fig. 5B). The ability of Salmonella spp. to utilize tetrahydronate as an electron acceptor during anaerobic respiration is a specific trait that is also used for the enrichment and identification of Salmonella spp. in clinical specimen. Tetrahydronate reduction is not required for systemic infection in a murine model, but may enable the colonization of certain anaerobic habitats by Salmonella spp.

SPI-1 is present in S. enterica as well as in S. bongori, while SPI-2 has been detected only in S. enterica; therefore, it is probably a more recent acquisition (147, 259). The acquisition of SPI-2 is considered an evolutionary step towards the systemic colonization of warm-blooded hosts (15a, 123). A common characteristic of both SPI-1 and SPI-2 is the observation that only a subset of the translocated effector proteins are encoded by genes within the island (Fig. 3A and 6). In fact, the majority of the effectors are encoded by distinct loci scattered over the chromosome (235, 239). Many of these loci are associated with bacteriophage genes, indicating that these elements have been acquired by separate events of horizontal gene transfer. SPI-1 and SPI-2 have evolved into stable regions of the Salmonella genome, and the loci each encode a T3SS and a small number of translocated effector proteins. The larger amount of effector proteins for either system, however, is encoded by separate loci outside of SPI-1 or SPI-2. The frequent association of the effector loci with cryptic, as well as functional, bacteriophages indicates that these effector genes encode a highly dynamic and mobile pool of virulence factors (91). The combination of effector genes in different serovars of Salmonella spp. may contribute to host specificity of the various serotypes as well as to disease outcome. This complex situation is schematically depicted in Fig. 6.

SPI-3. The genetic organization and function of SPI-3 are different from those of SPI-1 and SPI-2. This PAI is inserted at the selC tRNA gene, a locus that serves as insertion point for PAI in pathogenic E. coli as well (see the discussion of E. coli, below) (Table 3). The locus is about 17 kb, with an overall base composition similar to that of the core genome (20). There are several indications for a composite structure of SPI-3. Two fragments of IS elements were located in the central region of SPI-3, and the G+C content of genes within this locus is
### TABLE 3. Molecular and virulence characteristics of PAI of bacterial pathogens causing disease in humans, animals, and plants

<table>
<thead>
<tr>
<th>Species (pathogroup)</th>
<th>Strain’ or serovar</th>
<th>PAI name</th>
<th>Size (bp)</th>
<th>Insertion site</th>
<th>Stability</th>
<th>% G + C</th>
<th>Characteristic and virulence determinants encoded on PAI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong> (UPEC)</td>
<td>536</td>
<td>PAI I₃₃₆</td>
<td>76,843</td>
<td>tRNA selC</td>
<td>Low</td>
<td>46.0</td>
<td>Alpha-hemolysin, fimbrae, adhesins</td>
</tr>
<tr>
<td></td>
<td>536</td>
<td>PAI II₅₃₀</td>
<td>102,200</td>
<td>tRNA leuX</td>
<td>Low</td>
<td>46.0</td>
<td>Alpha-hemolysin, P-fimbriae, Hekadhesin, hemagglutinin-like adhesins</td>
</tr>
<tr>
<td>(UPEC)</td>
<td>536</td>
<td>PAI III₃₅₀</td>
<td>68,124</td>
<td>tRNA thrW</td>
<td>Low</td>
<td>47.0</td>
<td>S-fimbriae, tRNA siderophore system, Sap adhesin, TSH-like hemoglobin protease, HuMr-like heme receptor</td>
</tr>
<tr>
<td>(UPEC)</td>
<td>J96</td>
<td>PAI I₂₉₆</td>
<td>&gt;170,000</td>
<td>tRNA pheV</td>
<td>NP</td>
<td>NP</td>
<td>Alpha-hemolysin, P pilus</td>
</tr>
<tr>
<td>(UPEC)</td>
<td>J96</td>
<td>PAI II₅₉₆</td>
<td>&gt;110,000</td>
<td>tRNA pheU</td>
<td>NP</td>
<td>NP</td>
<td>Alpha-hemolysin, Pils (CNF1), heat-resistant hemagglutinin system</td>
</tr>
<tr>
<td>(UPEC)</td>
<td>CFT073</td>
<td>PAI CFT073</td>
<td>57,988</td>
<td>tRNA pheV</td>
<td>NP</td>
<td>42.9</td>
<td>Alpha-hemolysin, P-fimbriae, aerobactin</td>
</tr>
<tr>
<td>(UPEC)</td>
<td>CFT073</td>
<td>PAI II CFT073</td>
<td>&gt;71,684</td>
<td>tRNA pheU</td>
<td>NP</td>
<td>48.8²</td>
<td>P-fimbriae, iron-regulated genes</td>
</tr>
<tr>
<td>(MNEC)</td>
<td>C5</td>
<td>PAI I_C5</td>
<td>30,200</td>
<td>tRNA asnT</td>
<td>Low</td>
<td>57.0</td>
<td>Siderophore synthesis, iron uptake, see HPI</td>
</tr>
<tr>
<td>(MNEC)</td>
<td>RS218</td>
<td>GimA</td>
<td>~20,300</td>
<td>yjiA/DysE</td>
<td>NP</td>
<td>46.2²</td>
<td>Alpha-hemolysin, Pf, CNF1, heat-resistant hemagglutinin system</td>
</tr>
<tr>
<td>(SEPEC)</td>
<td>AL863</td>
<td>PAI I AL862</td>
<td>~61,000</td>
<td>tRNA pheR</td>
<td>NP</td>
<td>46.4²</td>
<td>Afa-EIII adhesin, ribose metabolism, PTS system</td>
</tr>
<tr>
<td>(EHEC)</td>
<td>EDL933</td>
<td>PAI II EDL933</td>
<td>43,590</td>
<td>tRNA selC</td>
<td>High</td>
<td>40.91</td>
<td>Translocation of T7, T3SS</td>
</tr>
<tr>
<td></td>
<td>LEE</td>
<td>LEE</td>
<td>68-24</td>
<td>tRNA selC</td>
<td>High</td>
<td>40.91</td>
<td>Iha (IrgA homolog adhesin), TlpA-D (tellurite³)</td>
</tr>
<tr>
<td></td>
<td>TAI</td>
<td>TAI</td>
<td>&gt;8,040</td>
<td>tRNA selC</td>
<td>High</td>
<td>40.91</td>
<td>Macrophage toxin, ClpB-like chaperone</td>
</tr>
<tr>
<td>(EHEC)</td>
<td>EDL933</td>
<td>OI #7</td>
<td>35,140</td>
<td>tRNA aspV</td>
<td>NP</td>
<td>NP</td>
<td>RTX-exotoxin, transport system</td>
</tr>
<tr>
<td></td>
<td>OI #28</td>
<td>OI #28</td>
<td>25,164</td>
<td>ybaT</td>
<td>NP</td>
<td>NP</td>
<td>Urease gene cluster</td>
</tr>
<tr>
<td></td>
<td>OI #43</td>
<td>OI #43</td>
<td>87,620</td>
<td>cdp</td>
<td>NP</td>
<td>NP</td>
<td>Adhesin, polypeptide biosynthesis system</td>
</tr>
<tr>
<td></td>
<td>OI #47</td>
<td>OI #47</td>
<td>31,726</td>
<td>ycaT</td>
<td>NP</td>
<td>NP</td>
<td>Urease gene cluster</td>
</tr>
<tr>
<td></td>
<td>OI #48</td>
<td>OI #48</td>
<td>87,547</td>
<td>ycfU</td>
<td>NP</td>
<td>NP</td>
<td>Urease gene cluster</td>
</tr>
<tr>
<td></td>
<td>OI #115</td>
<td>OI #115</td>
<td>16,947</td>
<td>Z4179</td>
<td>NP</td>
<td>NP</td>
<td>Urease gene cluster</td>
</tr>
<tr>
<td></td>
<td>OI #122</td>
<td>OI #122</td>
<td>23,454</td>
<td>pheV</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>OI #148</td>
<td>OI #148</td>
<td>43,417</td>
<td>tRNA selC</td>
<td>NP</td>
<td>NP</td>
<td>Contains TAI</td>
</tr>
<tr>
<td>(EHEC)</td>
<td>SPLE1</td>
<td>SPLE1</td>
<td>86,249</td>
<td>tRNA selC</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>SPLE2</td>
<td>SPLE2</td>
<td>13,459</td>
<td>yeeX</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>SPLE3</td>
<td>SPLE3</td>
<td>23,454</td>
<td>tRNA pheI</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>SPLE4</td>
<td>SPLE4</td>
<td>43,450</td>
<td>tRNA selC</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>SPLE5</td>
<td>SPLE5</td>
<td>10,235</td>
<td>tRNA leuX</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>SPLE6</td>
<td>SPLE6</td>
<td>34,148</td>
<td>tRNA leuX</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>(EPEC)</td>
<td>E2348/69</td>
<td>LEE</td>
<td>35,624</td>
<td>tRNA selC</td>
<td>High</td>
<td>38.36</td>
<td>Eae, translocation of T9, T3SS</td>
</tr>
<tr>
<td></td>
<td>E2348/69</td>
<td>EspC</td>
<td>15,195</td>
<td>ssaA</td>
<td>NP</td>
<td>40.5</td>
<td>EspC, ORF3</td>
</tr>
<tr>
<td>(EPEC)</td>
<td>RDEC-1</td>
<td>LEE</td>
<td>~34,000</td>
<td>liA</td>
<td>NP</td>
<td>NP</td>
<td>Eae, translocation of T4, T3SS</td>
</tr>
<tr>
<td>(EPEC)</td>
<td>RW1172</td>
<td>LEE</td>
<td>~80,000</td>
<td>tRNA pheV</td>
<td>NP</td>
<td>NP</td>
<td>Eae, translocation of T8, T3SS</td>
</tr>
<tr>
<td>(ETEC)</td>
<td>HT10407</td>
<td>Psl-T-PAI</td>
<td>46,000</td>
<td>tRNA selC</td>
<td>NP</td>
<td>NP</td>
<td>Tia, Leo</td>
</tr>
<tr>
<td>(ETEC)</td>
<td>Various</td>
<td>HPI</td>
<td>NP</td>
<td>tRNA amsT</td>
<td>NP</td>
<td>NP</td>
<td>Fya, Iri-1, 1, Ybt</td>
</tr>
<tr>
<td>(STECE)</td>
<td>4794/97</td>
<td>LPA</td>
<td>33,014</td>
<td>tRNA selC</td>
<td>NP</td>
<td>NP</td>
<td>47.4³</td>
</tr>
<tr>
<td></td>
<td>Various</td>
<td>HPI</td>
<td>NP</td>
<td>tRNA amsT</td>
<td>NP</td>
<td>NP</td>
<td>EspB, BubB, Iha</td>
</tr>
<tr>
<td></td>
<td>DBS1000</td>
<td>LEE</td>
<td>36,126</td>
<td>IS679/ABC²</td>
<td>NP</td>
<td>NP</td>
<td>Fya, Iri-1, 1, Ybt</td>
</tr>
<tr>
<td><strong>C. rodentium</strong></td>
<td>sv. Typhimurium</td>
<td>SPI-1</td>
<td>~40,000</td>
<td>fhuA</td>
<td>mid</td>
<td>42.0</td>
<td>Eae, translocation of T9, T3SS</td>
</tr>
<tr>
<td><strong>S. enterica</strong></td>
<td>sv. Typhimurium</td>
<td>SPI-2</td>
<td>~40,000</td>
<td>tRNA valV</td>
<td>High</td>
<td>44.3-55.5⁵</td>
<td>SeA, SeC, AbC, AbC, intracellular proliferation, T3SS</td>
</tr>
<tr>
<td></td>
<td>sv. Typhimurium</td>
<td>SPI-3</td>
<td>~17,000</td>
<td>tRNA selC</td>
<td>NP</td>
<td>47.5</td>
<td>MgIC, B, MarT, MsI (intracellular proliferation, Mg²⁺ uptake)</td>
</tr>
<tr>
<td></td>
<td>sv. Typhimurium</td>
<td>SPI-4</td>
<td>~25,000</td>
<td>ssh/ssSR</td>
<td>NP</td>
<td>37-54⁴</td>
<td>TISS, toxin secretion?, apoptosis</td>
</tr>
<tr>
<td></td>
<td>sv. Typhimurium</td>
<td>SPI-5</td>
<td>~7,000</td>
<td>tRNA erfT</td>
<td>NP</td>
<td>43.6</td>
<td>SopB (SigD), PBP (cofactor proteins for SPI-1 and SPI-2)</td>
</tr>
<tr>
<td></td>
<td>sv. Typhus</td>
<td>Major PAI/</td>
<td>143,000-</td>
<td>tRNA pheU</td>
<td>NP</td>
<td>NP</td>
<td>Vi antigen, SopE, type IV pilus</td>
</tr>
<tr>
<td></td>
<td>SPI-7</td>
<td>146,900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sv. Typhi CT18</td>
<td>SPI-6</td>
<td>~59,000</td>
<td>tRNA aop</td>
<td>NP</td>
<td>NP</td>
<td>safaA-D, tcaA-R usher-fimbrial operons</td>
</tr>
<tr>
<td></td>
<td>sv. Typhi CT18</td>
<td>SPI-8</td>
<td>~6,800</td>
<td>tRNA phe</td>
<td>NP</td>
<td>NP</td>
<td>Bacteriocin pseudogenes</td>
</tr>
<tr>
<td></td>
<td>sv. Typhi CT18</td>
<td>SPI-9</td>
<td>~16,000</td>
<td>ssaA</td>
<td>NP</td>
<td>NP</td>
<td>TISS, RTX-like protein</td>
</tr>
<tr>
<td></td>
<td>sv. Typhi CT18</td>
<td>SPI-10</td>
<td>~33,000</td>
<td>tRNA leu</td>
<td>NP</td>
<td>NP</td>
<td>safaA-R chaperone-usher-fimbrial operon</td>
</tr>
<tr>
<td>(DT104)</td>
<td>SGI</td>
<td>SGI</td>
<td>152,415</td>
<td>thdF</td>
<td>NP</td>
<td>41-58³</td>
<td>Multidrug-resistance, ampicillin² (Psee-1), chloramphenicol² (FloR, Cat), streptomycin² (AadA2), sulphonamides² (Sul I), and tetracycline² (TetR, TetG)</td>
</tr>
<tr>
<td>ssp. III, IV</td>
<td>HPI</td>
<td>HPI</td>
<td>NP</td>
<td>tRNA amsT</td>
<td>NP</td>
<td>NP</td>
<td>Fya, Iri-1, 1, Ybt</td>
</tr>
<tr>
<td><strong>Y. enterocolitica</strong></td>
<td>YE5081</td>
<td>HPI</td>
<td>~43,000</td>
<td>tRNA amsT</td>
<td>Low</td>
<td>60.0</td>
<td>Fya, Iri-1, 1, Ybt</td>
</tr>
</tbody>
</table>

Continued on following page
### TABLE 3—Continued

<table>
<thead>
<tr>
<th>Species (pathogroup)</th>
<th>Strain* or serovar</th>
<th>PAI name</th>
<th>Size (bp)</th>
<th>Insertion site</th>
<th>stability</th>
<th>Characteristic and virulence determinants encoded on PAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y. pseudotuberculosis</td>
<td>IP32637</td>
<td>HPI</td>
<td>~36,000</td>
<td>tRNA asuT, asuW</td>
<td>Mid 60</td>
<td>FyuA, YprI, ToxT (siderophore synthesis and uptake)</td>
</tr>
<tr>
<td>Y. pestis</td>
<td>6/69M</td>
<td>HPIgpm</td>
<td>119,443</td>
<td>tRNA asuT</td>
<td>Low 59</td>
<td>FyuA, YprI, ToxT, Ybt, pigmentation locus (siderophore synthesis and uptake)</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PAO1</td>
<td>PAGI-1</td>
<td>48,890</td>
<td>PAO2217</td>
<td>NP</td>
<td>54.9–63.7% Dehydrogenases, paraquat-inducible proteins, detoxification of reactive oxygen</td>
</tr>
<tr>
<td>S. flexneri</td>
<td>YSH6000T</td>
<td>SHI-1</td>
<td>46,603</td>
<td>tRNA pheV</td>
<td>Low 49.0</td>
<td>SigA, Pic (She), Shet1A and Shet1B, Sap (enterotoxin, proteases)</td>
</tr>
<tr>
<td>S. bovis</td>
<td>O1392</td>
<td>SHI-3</td>
<td>~21,000</td>
<td>tRNA pheU</td>
<td>NP</td>
<td>Aerobactin operon (iron uptake)</td>
</tr>
<tr>
<td>V. cholera</td>
<td>El Tor O1</td>
<td>VPI-1</td>
<td>~40,000</td>
<td>ssrA</td>
<td>NP</td>
<td>–35.0 Neuraminidase, utilization of amino sugars</td>
</tr>
<tr>
<td>N. gonorrhoeae</td>
<td>Various</td>
<td>elA</td>
<td>~8,900</td>
<td>NP</td>
<td>NP</td>
<td>41 NP all</td>
</tr>
<tr>
<td>S. enterica</td>
<td>50.5%</td>
<td>NP</td>
<td>~40,000</td>
<td>tRNA serV</td>
<td>NP</td>
<td>40.0 Deoxycytidylate deaminase-related protein, phospholipase A</td>
</tr>
<tr>
<td>L. monocytogenes</td>
<td>Various</td>
<td>LIPI-1</td>
<td>~9,000</td>
<td>NP</td>
<td>High NP</td>
<td>41 NP all</td>
</tr>
<tr>
<td>L. ivanovii</td>
<td>LIPI-2</td>
<td>~22,000</td>
<td>NP</td>
<td>Low NP</td>
<td>Low NP</td>
<td>41 NP all</td>
</tr>
<tr>
<td>S. aureus</td>
<td>Various</td>
<td>SCC mec</td>
<td>36,000–53,000</td>
<td>NP</td>
<td>NP</td>
<td>Methicillin (mecA), erythromycin (ermA), bleomycin (bleo), kanamycin (aadD)</td>
</tr>
<tr>
<td>E. faecalis</td>
<td>0100993</td>
<td>PPI-1</td>
<td>~27,200</td>
<td>OrfB</td>
<td>NP</td>
<td>32.6 PitABCD</td>
</tr>
<tr>
<td>E. coli</td>
<td>MMH594</td>
<td>NP**</td>
<td>153,571</td>
<td>H-RF</td>
<td>High NP</td>
<td>32.2 Cytolsyn (Cyl), biofilm formation (EspP)</td>
</tr>
<tr>
<td>C. difficile</td>
<td>VPI 10463</td>
<td>PaLoc</td>
<td>~19,600</td>
<td>cdu1, cdu2-2, cdd/2-4</td>
<td>High NP</td>
<td>Toxin A (TcdA), toxin B (TcdB)</td>
</tr>
</tbody>
</table>

*a Characteristic features of various PAI of bacterial pathogens are presented. tRNA indicates the presence of gene loci for tRNA that are adjacent to the PAI. Low, mid, and high stability indicate if the PAI is frequently deleted or stably inserted in the genome.

*b Indicates variation between different strains.

*c Strain that has been used for PAI identification and characterization.

*d The G+C contents of the entire chromosomal background (core) genomes are as follows: E. coli, ~50.5%; and S. enterica, 52 to 53%; V. cholerae El Tor, 47.7%; P. aeruginosa PAO1, ~66 to 67.0% Y. enterocolitica, 42.3%; Y. pseudotuberculosis and Y. pestis, 46 to 48%; S. pneumoniae, 40%; D. nodosus, 45%; H. pylori, 39%; N. meningitidis, 51.8%; N. gonorrhoeae, 50%; P. gingivalis, 48 to 57%; B. fragilis, 42%; E. faecalis, 57.5%; C. difficile, 26%; S. aureus, 32.8 to 32.9%.

*e G+C content in the sequenced region; may differ from the G+C content of the entire island.

*f Integrated with tandem with SPILE5.

*g Genes encoding an ABC transporter.

*h Similar to SaPI1.

*i Mobilizable by phage.

+j Rabbit-pathogenic E. coli.

+k Bovine isolate.

+l NP, not published.

+m NNP, no name published.

+n Mosaic structure.
alternating. Furthermore, SPI-3 shows heterogeneous structures in different subspecies of *Salmonella* spp. The main virulence factor encoded by SPI-3 is the high-affinity magnesium transport system MgtCB, which is important for the intracellular phenotype of *Salmonella* (19). For intracellular replication, *Salmonella* must adapt to the microbicidal and nutrient-poor environment of the phagosome, which is limiting for purines, pyrimidines, particular amino acids, and Mg\(^{2+}\). A large number of metabolic pathways and transport systems are required for adaptation to this environment (111). Mutant strains deficient in the MgtCB system are deficient in intracellular proliferation and systemic virulence. MgtB and MgtC are located in the cytoplasmic membrane. Whereas MgtB is a magnesium transporter (330), the function of MgtC is not yet clear (243). A further putative virulence gene is *misL*, encoding a putative T5SS (autotransporter) with similarity to VirG of *S. flexneri* and the AIDA-1 adhesin of EPEC (20). The role in virulence of *misL* and other genes within SPI-3 has not been elucidated.

**SPI-4.** By Southern hybridization analyses with probes of *E. coli* total DNA, Wong et al. (373) identified SPI-4 as a *Salmonella*-specific chromosomal region. This locus has characteristic features of a PAI and forms an insertion of 25 kb in *S. enterica* serovar Typhimurium but is less complex in *S. enterica* serovar Typhi (266). Putative virulence factors in SPI-4 are encoded by genes with sequence similarity to genes encoding T1SS for toxins and a gene that is required for the survival in macrophages (Table 3). The role of SPI-4 in *Salmonella* virulence has not been studied in full detail.

**SPI-5.** Of the loci encoding effector proteins for SPI-1 and SPI-2, SPI-5 is probably the most remarkable. The locus is only 7 kb long, has a G+C content of 43.6%, in contrast to 52% for the core genome of *S. enterica*, and is located adjacent to the *serF* tRNA gene. SPI-5 harbors the gene for SopD, an effector protein of the SPI-1-encoded T3SS (374), as well as *pipB*, encoding an effector protein for the SPI-2-encoded T3SS (188) (Table 3). This PAI is of composite nature since the portion that encodes an effector protein for the SPI-2-encoded system is not present in *S. bongori*, which also lacks the SPI-2 locus. SPI-5 was initially identified in *S. enterica* serovar Dublin but is conserved throughout *S. enterica*.

**Major PAI.** A PAI that is specific for *S. enterica* serovar Typhi and absent in other *S. enterica* subspecies I has been identified and named the major PAI (384). This locus is 146.9 kb long and is associated with the *pheU* tRNA gene (135). The major PAI contains several individually identified virulence factors: (i) the *viaB* gene locus for the biosynthesis of the Vi polysaccharide capsular of *S. enterica* serovar Typhi; (ii) the *sopE* prophage harboring the *sopE* gene encoding an effector protein of the SPI-1 system (240); and (iii) a gene cluster encoding type 4 pili with a function in invasion of epithelial cells by *S. enterica* serovar Typhi (385) (Table 3). For assembly of type 4 pili, subunits are secreted by the general secretion pathway, but the assembly of the pilus shaft occurs at a complex at the periplasmic site of the cytoplasmic membrane. An outer membrane pore is required for the extension of the growing pilus through the outer membrane. The major PAI harbors several genes associated with DNA mobility, indicating a mosaic structure of various horizontal acquisitions.

**SGI-1.** The emergence of multidrug-resistant strains such as DT104 is currently a major problem associated with *Salmonella* infections. Characterization of the resistance factors of such isolates led to the identification of a genomic island in multidrug-resistant strains of *S. enterica* serovars Typhimurium and Agona. This locus, termed *Salmonella* genomic island I (SGI-1), is 43 kb long and is flanked by DR. Further elements associated with DNA mobility, i.e., transposase, integrase, and excisionase genes with sequence similarities to transposon genes, have been detected (30). Within SGI-1, genes conferring the pentaresistance phenotype are clustered within the multidrug resistance region (Table 3). In contrast to plasmid-borne antibiotic resistance factors, the chromosomal SGI-1 appears to be stable in the absence of selective pressure. SGI-1 was also detected in *S. enterica* serovar Agona, indicating that transmission of the locus to other serovars may occur (30). It should be noted that multidrug resistance can also be mediated by plasmid-borne resistance genes, as observed by genome sequencing of *S. enterica* serovar Typhi harboring resistance plasmid pHCM1 (266).

**Other SPI.** The presence of further SPI has been inferred from the genome sequences of *S. enterica* serovars Typhi and Typhimurium (227, 266). Putative PAI in *S. enterica* serovar Typhi have been provisionally termed SPI-6 to SPI-10, with SPI-7 being identical to the major PAI. These additional loci have characteristics of PAI and contain genes of unknown function with sequence similarity to known virulence genes (Table 3).

**Enteropathogenic *E. coli***

**LEE of EPEC.** The locus of enterocyte effacement (LEE) was initially described in EPEC strains, causative agents of infant diarrhea in developing and industrialized countries (228). EPEC strains are able to cause attaching-and-effacing (A/E) lesions of the microvillus brush border of enterocytes; this ability was first described by Knutton et al. (190, 191), and later it was demonstrated that all the genes necessary for this phenotype are located on a PAI, termed LEE (82, 229). The LEE of EPEC reference strain E2348/69 is 35,624 bp long (82). Horizontal transfer of this element in vitro confers the whole A/E phenotype to laboratory *E. coli* strains (229). LEE contains 41 ORF and is organized in at least five polycistronic operons. It is integrated adjacent to either the *selC*, *pheU*, or *pheV* tRNA loci and consists of three functionally different modules. A T3SS exports effector molecules and is encoded at one end of the island. The secreted proteins EspA, EspB, EspD, and EspF, which function as part of the T3SS, are encoded at the opposite end of LEE. The central portion of LEE encodes intimin (Eae), which mediates intimate attachment to the host cell and Tir, the intimin receptor, which is chaperoned by CesT and translocated into the host cell plasma membrane by the T3SS (Table 3).

LEE genes are organized in five operons, and the regulation of LEE gene expression has been elucidated in detail (232). The components of the T3SS are transcribed from three operons, designated LEE1, LEE2, and LEE3. The secreted Esp effector proteins are transcribed from the polycistronic operon LEE4, and the fifth operon, *tir*, contains *tir*, *cesT*, and *eae*. The regulation of these operons is complex and involves the func-
tion of the LEE-encoded regulator (Ler), the plasmid-encoded regulator (Per), the DNA-binding protein H-NS, and integration host factor (IHF) (Fig. 3B). Furthermore, quorum sensing is also involved in that process. The plasmid-encoded regulator Per, encoded by perA, perB, and perC up-regulates the expression of the bundle-forming pili (BFP) and intimin in EPEC and regulates esp expression in response to environmental signals. Per also activates the LEE-encoded regulator Ler, which itself activates the expression of all LEE cistrons except LEE1. H-NS is a histone-like protein that shows similarity to Ler (81). It has been demonstrated that in the absence of H-NS, the expression of LEE2 and LEE3 became Ler independent. This observation led to the conclusion that Ler also acts as an antirepressor that antagonizes the H-NS-mediated silencing on the LEE2-LEE3 promoter region (39). Another level of LEE regulation involves the global regulator IHF (integration host factor), whose function is essential for the activation of ler (104). Quorum sensing also controls transcription of LEE1 and LEE2 operons, as shown by Sperandio et al. (333). LEE regulation represents an example of the complex network of PAI-encoded regulators and global regulatory systems that are encoded by the core genome. Examples of LEE regulation are schematically depicted in Fig. 3B.

The T3SS encoded by LEE is building a protein complex in the bacterial cell envelope; the complex is composed of Esp/ Sep proteins on which a filamentous structure of polymerized EspA is assembled (192). EspB and EspD are located at the distal end of the EspA filament, and it has been suggested that these proteins are involved in pore formation in the host cell membrane (138, 192, 196, 363, 366, 372). Effector proteins such as Tir, EspB, EspG, EspF, and Map are translocated through this structure (80, 182, 183, 230, 347). Translocation of these molecules into the host cells results in changes of the cytoskeleton of the underlying epithelial cells. Actin-rich pedestals are formed, and Tir is located at the apex of such pedestals, where it can interact directly with the bacterial outer membrane protein intimin. The N terminus of Tir is bound to the host cytoskeleton by α-actinin, while a phosphorylated tyrosine (Y474) of its C terminus binds to the host cytoskeletal adaptor protein (NCK) (for a review, see reference 380). NCK then recruits the N-WASP/ARP2/3 complex, which is involved in catalyzing the polymerization of actin beneath adherent bacteria. In addition, EPEC strains modulate other cellular processes such as transepithelial resistance and tight-junction permeability. Following tissue infection by EPEC, the myosin light chain and ezrin, proteins involved in the regulation of cellular permeability. Esp seems also to be involved in that process. The first stage is characterized by initial adherence via bundle-forming pili. The bacteria aggregate and are connected by bundles of BFP fibers. They then detach from the fibers, disaggregate, and become connected to the host cell through a surface appendage that contains EspA. It is thought that Tir, EspB, and EspF are transmitted via this appendage to the host cell. Tir and EspB are required for A/E activity. This results in pedestal formation and loss of the microvillus brush border. Tir localizes in the host cell membrane and interacts with intimin. Wiskott-Aldrich syndrome protein is then recruited to the pedestal, where it activates the Arp2-Arp3 complex to nucleate and polymerize actin.

The contribution of A/E lesions to diarrheal disease is not fully understood. Adherence of EPEC to polarized epithelial monolayers results in decreased transepithelial electrical resistance. In turn, this causes increased intestinal permeability (43). Increased permeability as a cause of decreased transepithelial electrical resistance has been hypothesized to be associated with diarrhea (335). Stein et al. (339) suggested that EPEC signaling leads to either an influx of positive ions or an efflux of negative ions across the membrane. Further experiments that are performed to explain the linkage of A/E lesions and diarrhea suggested that EPEC diarrhea could be the net result of multiple mechanisms. Early chloride secretion could lead to active secretory diarrhea, and the loss of epithelial cell microvilli leading to decreased absorption could account for persistent diarrhea, which is often a consequence of EPEC infection. Inflammatory mechanisms could also be involved in the onset of diarrheal symptoms (302).

EspC PAI. EspC is a 110-kDa protein secreted by EPEC strain E2348/69 that is not involved in A/E lesion formation and is not transported by the T3SS of the LEE (337). It belongs to the family of autotransporter proteins (TSSS) and shows serine protease activity. The espC gene is located in a 15,195-bp segment with a G+C content of 40.5%, which is integrated in the EPEC chromosome close to the iRNA-like gene ssrA, a site which is also used by the V. cholerae PAI VPI (Table 3). This espC PAI contains two genes, which potentially are associated with pathogenicity (espC and orf3), several ORF with similarities to mobile genetic elements, and an ORF putatively encoding regulatory proteins (233). EspC is homologous to the Pet autotransporter protein of enterogaegregative E. coli (EAEC), which has enterotoxic activity on intestinal tissue mounted in an Ussing chamber (87, 253). Furthermore, enterotoxic effects of supernatants of E. coli HBI01 containing EspC were observed by using this model (233). The protein encoded by ORF3 is homologous to VirA of S. flexneri and rORF2 of EPEC. The ability to invade cultured epithelial cells of the virA deletion strain of S. flexneri could be rescued by cloned ORF3 or rORF2. However, an ORF3-rORF2 double mutation did not affect EPEC invasion; therefore, the function of these proteins remains obscure. The EspC island is found only in a subset of EPEC strains. EspC most probably plays an accessory role in EPEC pathogenesis as an enterotoxin. This, however, should be investigated in vivo (233).

LEE of rabbit-pathogenic E. coli. Particular E. coli strains are associated with diarrhea and other enteric infections in rabbits, pigs, calves, lambs, and dogs (17, 44, 166). It was shown by Zhu et al. (387) that rabbit diarrheagenic E. coli strain RDEC-1 contains a LEE that is not inserted at selC and is...
flanked by an IS 2 element and the lifA toxin gene. The RDEC-1 LEE contains a core region sharing 40 ORF with EHEC LEE and EPEC LEE. However, ORF3 and ERIC sequences are absent. The overall genetic order, as well as the sequence of ese, sep, cex, and several ORF and RORF genes, is consistent with their conserved function. Genes whose products interact with the host, some regulatory regions, and genes flanking the LEE are divergent from those of EPEC LEE and EHEC LEE (387). These variations have been suggested to be important for adaptation to a specific pathogenic life-style. In a further study, it was demonstrated that LEE of rabbit-specific E. coli strains was deleted spontaneously and that a cloned phe-specific integrase including the attR site of the attP site of LEE mediated site-specific integration at the pheU locus. This was taken as support for phage-mediated transfer of LEE (346). LEE has also been characterized in a bovine Shiga toxin-producing E. coli (STEC) O103:H2 strain. In contrast to LEE of EPEC and EHEC, this LEE contains DNA in addition to the 35-kb core region, so that its size is at least 80 kb (170).

Enterohemorrhagic E. coli

EHEC strains are a subgroup of STEC. These pathogens can cause a spectrum of disease ranging from uncomplicated watery diarrhea to bloody diarrhea with abdominal cramps. EHEC-associated bloody diarrhea is considered to be a risk factor for the development of the hemolytic-uremic syndrome, which is defined by the triad of thrombocytopenia, hemolytic anemia, and renal insufficiency (119, 120). EHEC pathogenicity is determined mainly by pathogenicity factors encoded by mobile genetic elements, the most cardinal factor of which is the phage-encoded Shiga toxin (Stx) (119, 120). Besides the production of Stx, EHEC strains harbor a large plasmid encoding various putative virulence factors such as the enterohemolysin E-Hly (306, 308), the serine protease EspP (37), the catalase-peroxidase KatP (36), the ToxB protein with homology to large clostridial cytotoxins (345), and the StcE metallo-protease which is transported by the etp T2SS into the extracellular environment (206, 307). Following the sequencing of the E. coli O157:H7 genome, other putative virulence factors have been discovered (271).

LEE of EHEC. EHEC strains also possess a LEE, first detected and sequenced in EDL933 (83, 228, 270, 332). Proteins encoded by the LEE of EHEC are also involved in the A/E phenotype, and this locus is inserted in the same sites as the LEE of EPEC strain E2348/69 (for a review, see reference 251). The LEE of EHEC contains 36 ORFs, of which 34 are common to the LEE of both EPEC and EHEC. The remaining 13 ORFs belong to a putative prophage, designated 933L, which is located close to the selC locus. The difference in the sizes of the EHEC and EPEC LEE, i.e., 43,359 and 35,624 bp, respectively, is therefore due mainly to insertion of this P4-like prophage element (270). Although LEE-encoded proteins of EHEC and EPEC show an overall sequence identity of 93.9%, some gene products, such as EspB (74% identity), EspD (80.4%), and EspA (84.6%), show higher sequence diversity. Other proteins that differ markedly are intimin (87.7%) and Tir (66.5%). In EHEC O157:H7 strains, Tir is not phosphorylated on translocation in the eukaryotic host cell. This is due to the change of a serine to a tyrosine (180, 181). In contrast to the LEE of EPEC, cloned EHEC LEE is not able to confer A/E lesions to an E. coli laboratory strain. It has been suggested that the different phenotypes produced by LEE of EPEC and EHEC are based on natural selection for adaptation to the host microenvironment or to evasion of the host immune system (270).

Other genetic islands. The sequencing of the E. coli O157:H7 genome was performed independently on two E. coli O157:H7 strains by two groups in the United States and Japan, and the genomes were published in 2001 (200, 271). Although the two genomes have numerous similarities, they differ in a couple of features. Analysis of the EDL933 genome has led to the discovery of DNA regions designated O-islands that are present in E. coli O157:H7 strain EDL933 but not in E. coli K-12 and regions designated K-islands that are present in E. coli K-12 but not in EDL933 (271). Besides a common colinear backbone of 4.1 Mb, 1.34 Mb of DNA is located in O-islands and 0.53 Mb is located in K-islands.

Roughly 26% of the genes of strain EDL933 are located in O-islands. A total of 177 O-islands have been detected that are larger than 50 bp, and 9 of these are larger than 15 kb and were found to encode putative virulence factors. One of these, OI#148 encodes LEE, and the other eight encode a macrophage toxin and ClpB-like chaperone (OI#7), an RTX-like exoprotein and a transport system (OI#28), two urease gene clusters (OI#43 and OI#48), an adhesin and polyketide or fatty acid biosynthesis system, (OI#47), a T3SS and secreted proteins similar to the Salmonella-Shigella host cell invasion genes of SPI-1 (OI#115), two toxins and a PagC-like virulence factor (OI#122), and a fatty acid biosynthesis system (OI#138) (Table 3). There are also smaller islands that harbor virulence-associated genes such as fimbrial biosynthesis genes, iron uptake and utilization clusters, and putative nonfimbrial adhesions, as well as putative systems for carbohydrate transport, antibiotic efflux, aromatic compound degradation, tellurite resistance, and glutamate fermentation.

Except for OI#148, encoding LEE, these factors have not been linked directly to the pathogenicity of EHEC O157, and it has been suggested that they are putative factors for survival in the environment and colonization in the human host (271). Hayashi et al. (140) termed the regions specific for E. coli O157:H7 strain RIMD0509952 S-loops (strain-specific loops) and the K-12-specific DNA K-loops. These authors found 18 prophages and 6 prophage-like elements (elements without homology to prophages except for integrase genes, which are termed SPLE for “Sakai prophage-like elements”). SPLE1 is 82.2 kb long, and part of it has been designated “tellurite resistance and adherence-conferring island” (TAI; see below). SPLE4 includes the LEE locus (Table 3). The term “SPLE” is apparently been used synonymously with PAI, and a number of other virulence-related genes were described to be located in these mobile genetic elements (140).

TAI. An outer membrane protein of E. coli O157:H7, designated Iha, could be demonstrated to mediate adherence to HeLa cells after transformation of a recombinant clone to a nonadherent laboratory E. coli strain (344). Iha is homologous to various bacterial iron acquisition proteins but not to other adhesins. The iha gene is present on a chromosomal 8,040-bp KpnI fragment located close to four additional genes, which are homologous to tellurite resistance gene of Alcaligenes spp.
and Serratia marcescens. They were designated tlpa to tlpd (for “tellurite resistance proteins”) (Table 3). The role of this island in EHEC pathogenicity should be further evaluated (344). The deduced Iha protein is 53% identical to IrgA of V. cholerae, which is encoded by the iron-regulated gene irgA. The 8,040-bp KpnI fragment was part of a 35-kb cosmid clone, whose insert was designated TAI (see above). TAI is dispersed in E. coli O157 isolates, and certainly Iha was found in various non-O157 STEC strains (309). The boundaries of this putative island and its role in the pathogenesis of diseases associated with O157:H7 are not yet clearly established. TAI sequences have been identified in SPE1 in the genome sequence of E. coli O157:H7 strain RIMD0509952 (see above).

Other Intestinal E. coli Groups

HPI of pathogenic E. coli. The high-pathogenicity island (HPI) was first described in pathogenic Yersinia spp. and is described in detail in the Yersinia section (see above). However, HPI is not restricted to Yersinia. It has also been detected in enteroaggregative E. coli (EAEC) as well as in enteroinvasive (EIEC), enterotoxic (ETEC), enteropathogenic (EPEC), and Shiga toxin-producing (STEC) E. coli, and extraintestinal E. coli, and Salmonella enterica strains (Table 2) (175, 264, 311).

HPI has been proposed to contribute to the fitness of E. coli under certain environmental conditions and is one of the most widespread PAI so far detected, being present with some modifications in a variety of enterobacteria. The integration of HPI into one of three asnT tRNA genes is a feature of all HPI variants and is probably caused by an integrase specific for asnT (312). The presence of HPI in bacteria using different pathogenic strategies supports the theory that it is not involved in pathogenesis but in increasing pathogenic strategies supports the theory that it is not involved (312). The presence of HPI in bacteria using different pathogenic strategies supports the theory that it is not involved in pathogenesis but in increasing pathogenic strategies supports the theory that it is not involved (312). The presence of HPI in bacteria using different pathogenic strategies supports the theory that it is not involved in pathogenesis but in increasing pathogenic strategies supports the theory that it is not involved.

The widespread presence of HPI in different species and pathotypes also implies an efficient mechanism of transfer. Therefore, HPI could be a good model for studying the transfer and function of HPI and will give detailed insights into the biology and evolution of PAI.

LPA of Stx2d-producing E. coli. STEC strains contain genes encoding proteins of the major Stx groups, Stx1 and Stx2. The Stx2 group is a family of structurally related proteins, including Stx2 and its variants, designated Stx2c, Stx2d, Stx2e, and Stx2f. Recently, Stx1 variants have been described (105, 234, 383).

In a recent study, the locus of proteolysis activity (LPA), a 33,014-bp genomic island, which is integrated downstream of selC, was identified in Stx2d-producing eae-negative E. coli strains. LPA has a G+C content of 47.4% and contains 24 ORF. Some of them are thought to be important for interaction with the eukaryotic host. One of these is EspI, a protein that belongs to the autotransporter (TSSS) class of proteins. EspI is a serine protease, which is able to cleave human coagulation factor V. The second putative virulence gene is iha, encoding an adherence factor that was previously described for E. coli O157:H7 (see “Enterohemorrhagic E. coli” above). Whether the observed in vitro activities of EspI and Iha contribute to disease or represent only an environmental advantage has yet to be shown (309).

Enterotoxigenic E. coli

ETEC strains are a major cause of diarrheal disease in childhood in developing countries and a frequent cause of traveler’s diarrhea among visitors to these countries (251). They are E. coli strains that produce at least one member of the heat-stable or heat-labile groups of enterotoxins. In addition, they produce colonization factor antigens necessary for binding to the small-intestine mucosa. The initial observation that ETEC strains could invade human intestinal cell lines in vivo led to the discovery and characterization of two chromosomal loci, tia and tib (toxigenic invasion loci A and B) of ETEC strain H10407, which were found to be associated with adherence to and invasion of intestinal epithelial cells (84). The tia locus is involved in the synthesis of Tia, a 25-kDa outer membrane protein, obviously binding to epithelial cells in a receptor-mediated manner and causing the adherence and invasion phenotype (94). The tia locus is located on a large chromosomal element of ca. 46 kb, which is inserted at selC tRNA locus (95). The Tia PAI contains a P4-like phase integrase gene and has a G+C content of 43.7% (Table 3). In addition to tia, four further candidate ORF encoding a bacterial secretion apparatus were identified. One of these genes is needed for maximum secretion of the LT enterotoxin. This gene was designated leaA, for “labile enterotoxin output” (95). Another ETEC protein, TibA, is a 104-kDa outer membrane protein thought to be involved in adherence and invasion. It could be shown that purified TibA protein specifically binds to cultured human intestinal cells and that polyclonal TibA antisera decreases TibA-mediated bacterial invasion of epithelial cells (211). The functional form of TibA is generated by glycosylation of Pre-TibA, a process highly reminiscent of the activation of AIDA-I, an adhesin of diffusely adhering E. coli (DAEC) (85). This modification is mediated by the TibC protein, a putative heptosyltransferase (J. G. Mammarappallil, B. W. Hronek, L. D. Mettenburg, and E. A. Elsinghorst, Abstr. 101st Gen. Meet. Am. Soc. Microbiol. 2001, abstr. B57, 2001). The tib locus appears also to be a chromosomal island distinct from tia; however, more information is needed to prove this assumption.

Enteroaggregative E. coli

EAEC strains confer a specific phenotype upon adherence to HEP-2 cells. They attach in a stacked brick-like pattern that is designated aggregative adherence. EAEC strains have been associated with persistent diarrhea in young children in developing countries, with traveler’s diarrhea, and with sporadic endemic diarrhea worldwide (251, 252). Diarrhea caused by EAEC is typically nonbloody, and the bacteria do not invade epithelial cells and do not cause inflammation. Most of the virulence factors associated with EAEC-related disease are encoded by a large plasmid. These factors included the fimbrial antigens AAF/I and AAF/II and the putative enterotoxins Pet and EAST1 (87, 301). However, some chromosomal loci associated with virulence have been identified and linked to puta-
tive PAI. One of these represents a set/pic locus, similar to SHI-1 of S. flexneri; another is HPI (18, 324, 350).

**Enteroinvasive E. coli**

EIEC strains are closely related to Shigella spp. and resemble them in their pathogenic mechanisms as well as in the diseases they cause. EIEC strains also contain aerobactin genes; however, these are not located in SHI-2-like islands. It could be shown by hybridization experiments that EIEC aerobactin genes are associated with the selC tRNA locus and probably are located in a novel PAI, but further studies are needed to gain more insight in the genetic background of these genes (350, 361). A fyuA-irp gene cluster is infrequently found in EIEC, indicating the presence of HPI (313).

**Uropathogenic E. coli**

Urinary tract infections (UTI) are some of the most common nosocomial infections, and E. coli is the most prevalent bacterial causative agent of this disease. The pathogenic mechanisms underlying this disease have already been investigated in detail. UPEC strains initiate infection by binding to the superficial bladder epithelial cells that line the luminal surface of the bladder. In the majority of cases, this is achieved by type 1 pili. This adherence prevents the pathogen from being washed out by the urine flow. Subsequently, bladder cells internalize the UPEC, a process that is considered an escape mechanism to protect the bacteria from the host immune system. However, internalization causes exfoliation of the superficial bladder cells harboring internalized bacteria into the urine. Before clearance, intracellular bacteria start to replicate and induce host cytokine responses, leading to the symptoms of UTI (221, 246, 305).

UPEC strains express a collection of strain-specific antihost weapons. UPEC strains 536, J96, and CFT073 have been characterized in detail (288). Most UPEC strains express type 1 pili in vivo, which play a role in internalization. Furthermore, Dr adhesins have been described to mediate internalization via α4β1-integrin. Also, P-fimbriae might play a role in the enhancement of bacterial colonization in human UTI.

UPEC strains express an array of virulence factors, which are often encoded on PAI. Such virulence factors, which are associated with UTI, include the toxins cytotoxic necrotizing factor 1 (CNF1), which has been demonstrated to increase F-actin polymerization, superoxide generation, and adherence to T84 cells (154), and alpha-hemolysin. CNF1, encoded by PAI II36 of UPEC, belongs to the family of Rho-GTPase-activating bacterial toxins. CNF1 is a 108-kDa protein toxin containing three domains which are implicated in the three steps of the intoxication process. The N-terminal domain binds with high affinity to an unknown cell receptor. Binding of the toxin is followed by its internalization by endocytosis and its transport into late endosomes. The central domain of the toxin contains two hydophobic helices, which allow translocation of the toxin across the membrane on acidification in late endosomes. Finally, the carboxy-terminal domain of CNF1 is an enzyme that deaminates Rho-GTP-binding proteins (Rho, Rac, and Cdc42) at glutamine 63 (for Rho) or glutamine 61 (for Rac and Cdc42). Activation of Rho GTPases by CNF1 induces a profound reorganization of the cell actin cytoskeleton (29, 310). Eukaryotic cells intoxicated with CNF1 exhibit membrane ruffling, formation of focal adhesions and actin stress fibers, and DNA replication in the absence of cell division, a phenomenon that results in enlarged, multineucleated cells. In vivo, CNF1 causes necrosis in rabbit skin following intradermal injection and also causes persistent inflammation in a mouse footpad assay. Epidemiologic data support the role of CNF1 as a virulence factor in human extraintestinal infections, although direct proof of the role of this toxin in disease remains to be found (for a review, see reference 310).

Alpha-hemolysin belongs to the Repeat in Toxin (RTX) group of bacterial toxins with pore-forming activities in the membrane of erythrocytes and other eukaryotic cells. Some of these toxins are translocated by a TISS through the bacterial cytoplasmic and outer membranes. The alpha-hemolysin operon of UPEC can be located on either a plasmid or the chromosome. Four genes, hlyC, hlyA, hlyB, and hlyD, determine the synthesis, activation and transport of alpha-hemolysin outside the bacterial cell (for a review, see reference 368).

E. coli hemolysin targets erythrocytes but also mammalian cells such as endothelial cells and renal epithelial cells, as well as immune cells such as granulocytes and monocytes. At high concentrations, hemolysin generates cation-selective pores in eukaryotic cell membranes and causes target cell lysis. Lower levels of hemolysin trigger G-protein-dependent reactions and thus stimulate the respiratory burst and the secretion of vesicular constituents. Early work performed with a pyleonephritis mouse model demonstrated that alpha-hemolysin enhances nephropathogenicity (362). In addition, medication of cell death via lymphocyte function-associated antigen (Lfa-1) has been shown for alpha-hemolysin. Therefore, this toxin is involved in triggering apoptosis (195, 367). It has also been suggested that alpha-hemolysins are involved in renal failure in patients with septic shock due to gram-negative bacteria UPEC strains also produce the siderophores aerobactin and enterobactin, LPS, capsules, and a number of adhesive molecules (246, 263). The model organism for the study of PAI in extraintestinal E. coli is the uropathogenic isolate 536 (O6:K15: H31). Four PAI have been identified in the chromosome of this organism and have been named PAI 536 to PAI IV 536 (Table 3).

**PAI I536 to PAI IV536.** PAI I536 is associated with the selC tRNA gene and is 76,843 bp long. It is flanked by 16-bp DRs and has a G+C content of 46%. A nonfunctional phage P4-like integrase gene is located directly downstream of selC. Besides the well-known virulence factors alpha-hemolysin, F17-like fimbriae, and CS12-like fimbriae, a couple of genes for yet unidentified putative fimbriae, adhesion determinants, and an ORF whose product showed homology to modification methylase NgoVII were discovered.

**PAI IV536** is associated with the leuX tRNA gene and is 102,200 bp long. The G+C content is also 46%, and the integrase gene downstream of leuX is functional and highly homologous to the phage P4 integrase gene. PAI IV536 contains genes for P-related fimbriae, another alpha-hemolysin gene cluster, the Hek adhesin, hemagglutinin-like adhesins, and an ORF whose product shows similarity to a fragment of the HgiDII modification methylase of Herpetosiphon aurantiacus (68).
PAI III536 is integrated downstream of the thrW tRNA gene, is 68,124 bp long, and has a G+C content of 47%. This PAI begins with an integrase gene homologous to that of phage Sfx, which recognizes this tRNA gene as an integration site. PAI III536 encodes S-fimbriae, the iro siderophore system, a HmuR-like heme receptor, a Sap adhesin, and a TSH-like hemoglobin protease (68).

PAI IV536 is associated with the ounT tRNA gene. PAIV536 represents the core element of the HPI of pathogenic Yersinia spp. It is 30.2 kb long and encodes the yersiniabactin siderophore system (68).

The occurrence and distribution of these PAI in extraintestinal E. coli has been investigated, and it could be shown that not all PAI were present in each of the UTI isolates, meningitis E. coli (MNEC), sepsis E. coli (SEPEC), and diarrheagenic isolates tested. PAI IV536 had the highest prevalence among these isolates, and it was also shown that even similar PAI differ in their size, gene composition, and sequence among the various strains.

PAI IJ96 and PAI IIJ96. Up to now, two PAI, PAI IJ96 and PAI IIJ96, have been found in UPEC strain J96. PAI IJ96 is more than 170 kb long and contains a hly operon and a pap operon encoding a P-pilus. The assembly of the P-pilus and that of type 1 pili are the prototypes for a process mediated by the usher-chaperone pathway. Fimbrial subunits are secreted by the general secretion pathway into the periplasm. The subunits associate with a periplasmic chaperone protein and are transported to an usher in the outer membrane. The growing pilus is extended by addition of subunits at the periplasmic side of the outer membrane. The binding specificity of the pilus is determined by protein subunits located at the tip of the pilus, and structural analysis of tip proteins revealed the nature of molecular interactions with host molecules (for a review, see reference 300).

PAI IIJ96 is 110 kb long and also contains a hly operon, as well as a prs operon encoding a different class of pilus and a gene encoding CNF-1: Whereas PAI IJ96 is integrated adjacent to pheV, PAI IIJ96 is integrated downstream of the pheU gene (25, 288, 343, 369).

PAI ICF073 and PAI II CF073. UPEC strain CF073 was isolated from the blood and urine of a woman with pyelonephritis. Genetic analysis of CF073 led to the discovery of PAI ICF073, which carries an alpha-hemolysin operon (hly), a pap operon encoding P-fimbriae, and genes related to iron transport systems and putative carbohydrate transport systems (241, 369). PAI ICF073 is 58 kb long, has a G+C content of 42.9%, and comprises 44 ORF and segments of E. coli K-12 core chromosome interspersed with this PAI (Table 3). From PAI II CF073, 71,684 bp has been sequenced without reaching the borders between the PAI and the core chromosome. The sequenced portion comprises 89 ORF and contains a pap operon, iron-regulated genes, mobile genetic elements, and a large number of unknown genes. Eleven probes were prepared from this island for hybridization studies to clinical isolates. The results demonstrated that seven sequences were more prevalent among uropathogens, two regions were more prevalent among cystitis and pyelonephritis isolates, two were prevalent among pyelonephritis isolates only, and three were prevalent among cystitis isolates only. The authors suggested that groups of pathogens have genetic differences that may be responsible for the different clinical outcomes (287).

Other Extraintestinal E. coli Strains

E. coli is also responsible for one-third of cases of neonatal meningitis, with an incidence of 0.1 per 1,000 live births in the United States (66). Only a few virulence determinants specific for E. coli causing meningitis (MNEC) have been described. The expression of the capsular polysaccharide K1 (194) and the siderophore aerobactin (254) are thought to be important for bloodstream dissemination. S-fimbrial adhesin (sfa) and Ibe10 protein are involved in adhesion and invasion of brain microvascular endothelial cells, and they are likely to promote the crossing of the blood-brain barrier (130, 161, 268). Recently, traI, cnf1, ciga, and nla to nilC have been incriminated in MNEC virulence (12, 13, 153, 160, 161). However, not all meningitis isolates express all these factors, and it has been hypothesized that others must be present and involved in the two major pathophysiological steps. Houdouin et al. (157) reported that genes encoding P-related fimbriae (prs), hemolysin (hly), CNF (cnf), and a heat-resistant agglutinin (hra) were found clustered in the archetypical MNEC strain C5. They provided evidence that this gene cluster constitutes a PAI, similar to PAI IJ96 of UPEC strains. In contrast to E. coli J96, in E. coli C5 this cluster is inserted in the leuX tRNA gene. The authors demonstrated with a rat model that this PAI contributes to survival of the bacteria in the bloodstream and is therefore most probably involved in the bacteremic phase of neonatal meningitis. A role for this PAI, which has been designated PAI Ic5, in the development of meningitis or resistance to human serum could not be shown.

In a recent publication, Huang et al. (159) reported that E. coli O18:K1:H7 strain RS218, isolated from the cerebrospinal fluid of a newborn baby, contained a genetic island (GimA). GimA harbors a gene encoding IbeA, an invasion determinant that is involved in crossing of the blood-brain barrier. GimA is 20.3 kb long, has a G+C content of 46.2%, and contains, in addition to IbeA, 14 genes, mainly encoding carbon source metabolism and substrate transport. In their study, the authors showed that glucose enhances the penetration of human brain vascular endothelial cells by E. coli and that cyclic AMP was able to block the stimulating effects of glucose. It was suggested that this island contributes to invasion of the blood-brain barrier by E. coli through a carbon source-regulated process.

E. coli strain AL863, a blood isolate from a cancer patient, was demonstrated to carry two genomic islands at the pheV and pheR tRNA loci. However, except for integration sites, these islands have only low concordance with other extraintestinal E. coli islands (201). PAI IAL862 is 61 kb long with a G+C content of 46.4%. The AfaE-VIII adhesin, encoded by the afa-8 operon, is the only known virulence factor encoded by this island. However, gene clusters putatively encoding sugar utilization pathways have been discovered, suggesting a more ecological role of this island. PAI II AL862 contains a P4-like integrase gene and 14-bp DRs at both ends. PAI II AL862 is also 61 kb long and resembles PAI I. It also contains the afa-8 gene cluster. However, the islands are yet not completely sequenced, and the information is therefore only partial. It is
thought that either PAI II\textsubscript{AL862} is a duplication of PAI I or that two similar isolates have been acquired independently. The fimbrial adhesin AfaE-VIII is produced by animal-pathogenic isolates associated with diarrhea and septicemia and by human isolates associated with extraintestinal infections.

**Pathogenic Neisseria spp.**

The Gram-negative diplococci *Neisseria meningitidis* and *N. gonorrhoeae* (here referred to as pathogenic *Neisseria*) are causative agents of serious human diseases. Infection with *N. meningitidis* can cause a spectrum of clinical syndromes including bacterial meningitis and sepsis, respiratory tract infections, and meningococcemia (purpura fulminans and Waterhouse-Friderichsen syndrome). Much less frequently, other syndromes are associated with meningococcal disease, such as conjunctivitis, otitis media, epiglottitis, arthritis, urethritis, and meningococcemia (purpura fulminans and Waterhouse-Friderichsen syndrome). Each of these syndromes is characterized by the presence of meningococci (malignant bacterial meningitis and sepsis, respiratory tract infections, and meningococcemia) and meningococcemia (purpura fulminans and Waterhouse-Friderichsen syndrome). Much less frequently, other syndromes are associated with meningococcal disease, such as conjunctivitis, otitis media, epiglottitis, arthritis, urethritis, and meningococcemia (purpura fulminans and Waterhouse-Friderichsen syndrome).

Meningococci are a highly diverse species capable of effectively exchanging genetic material and thereby switching serotypes (359, 360). The human nasopharynx is the natural reservoir of *N. meningitidis* and the site from which infections start. Meningococci overcome host defenses and attach to the microvillus surface of non-ciliated columnar mucosal cells of the nasopharynx. Binding stimulates engulfment of the meningococci by epithelial cells, which may then traverse the mucosal epithelium through phagocytic vacuoles. In a small number of persons, meningococci may then penetrate the mucosa and gain access to the bloodstream, causing systemic disease; moreover, they may cross the blood-brain barrier to cause meningitis. Important pathogenicity factors of meningococci are the capsule, pili, iron-acquisition factors, Opa proteins, porins, and the release of endotoxin (250).

*N. gonorrhoeae* is the causative agent of the sexually transmitted disease gonorrhea. The main clinical presentations of uncomplicated gonorrhea are urethritis in men and endocervicitis in women. Most gonococcal infections remain localized to the genital tract, but in a small proportion of untreated cases, a systemic infection develops and causes the serious complication of disseminated gonococcal infection (DGI). DGI occurs in 0.1 to 0.3% of patients and is usually characterized by a severe arthritic condition and can also produce tenosynovitis, dermatitis, meningitis or endocarditis (79, 258). Clinical manifestations of endotoxin (250).

Analysis of an 8.9-kb DNA segment flanking *atlA* revealed a G+C content of 43%, in contrast to 50% for the rest of the chromosome. This DNA fragment contains only one gonococcal transformation uptake sequence, whereas this sequence occurs every 2 kb in the majority of the chromosome. In addition to *atlA*, two ORF similar to *traG* and *traH* of *E. coli*, an ORF encoding a cold shock protein homologue (CspA), and ORF *exp1* and *exp2* have been detected. Whereas *exp1* encodes an exported protein, the function of *exp2* has not been determined. The function of ORF7 and ORF8 is also not known. Genetic experiments suggest that the *atlA* region is part of 60- to 70-kb variable genetic island in a subset of gonococcal isolates. This genetic island is heterogenic in different gonococcal isolates, but it has been shown that it does not occur in meningococci (67).

The *sac-4* locus that contains an uncharacterized DNA fragment mediating serum resistance has recently been shown to be an allele of the *traG* gene. It was shown that this island is highly variable and that different gene compositions occur in different gonococcal strains. The presence of *atlA* and *sac-4* is statistically significantly associated with DGI. It has been suggested that the *atlA* island encodes a DNA secretion system and that the *tra* genes are part of that system.

The presence of the serum resistance allele *sac-4* in DGI isolates is consistent with an ability to survive in the bloodstream to cause disseminated disease. The expression of AtlA may increase the level of toxic PG fragments to modulate the immune response, exacerbate DGI arthritis, and potentiate meningitis. There are, however, some observations that are not consistent with this model, and future research is needed to clarify this island role in pathogenesis.

One of the remarkable features of pathogenic *Neisseria* species is that these organisms are naturally transformable and that the neisserial genome has undergone a rapid evolution by high-frequency recombination and DNA uptake (155). Because of this high rate of microevolution, the emergence of new clonal groupings responsible for outbreaks of meningococcal disease can be readily detected by genetic typing methods (47). Transformation and recombination of foreign DNA into the *Neisseria* chromosome could also facilitate the acquisition of new genes. DNA fragments containing uptake sequences are preferentially incorporated in the genomes of *N. meningitidis* and *N. gonorrhoeae*. Various genetic islands or DNA cassettes have been proposed to be acquired by transformation and recombination, such as a DNA island containing the capsule biosynthesis locus and the gene cassettes for glutathione peroxidase, rotamase, and the RTX-like Frp proteins. At least eight genetic islands have been identified in meningococci by subtractive hybridization techniques (185).

An exchangeable DNA island (*exl*) consisting of multiple cassettes that are capable of being exchanged for the *hmbR* gene in the meningococcal genome has been detected (172). A phase-variable hemoglobin receptor is encoded by *hmbR*, which is suggested to be necessary for iron uptake. The *hmbR* locus (*excl1*) is located between *hemO* and *col* in *N. meningitidis* strain NMB and is exchanged with *excl2* or *excl3* in different meningitis isolates. The *excl1* island also occurs in *N. gonorrhoeae* strain FA 1090 and contains four elements. Downstream of *col*, there is a 524-bp ORF, designated ORFU, which is flanked by two 33-bp repeats containing neisserial
DNA uptake sequences. In *N. meningitidis* serogroup N strain NMB, ORFU is replaced by a 1,280-bp sequence containing ORFX and ORFY. In different strains these sites occur in various combination, suggesting a high level of recombination. Serogroup Y strain GA0290 and serogroup W-135 strain 6083 do not contain excl1 (*hmbR*). The 1,904-bp cassette in strain GA0929 contains the 809-bp ORF excl2 instead of *hmbR*. Strain 6083 contains a 3,426-bp insertion carrying the 1.368 ORF excl3. The deduced amino acid sequences of the excl2 and excl3 products are 55% identical and have low sequence similarity to putative lipoproteins occurring in meningococci. The excl loci have also been found in commensal *Neisseria* strains (172).

The ability of pathogenic *Neisseria* species to establish disease is connected to their ability to scavenge essential nutrients, such as iron, in the human environment. Pathogenic *Neisseria* species express, instead of siderophore systems, at least four outer membrane protein receptors that are specific for transferring (*bhpAB*), lactoferrin (*lbpAB*), hemoglobin (*hmbR*), and Hb-Hp-complexes (*hpuAB*). After being translocated into the cytoplasm, iron is released by a heme oxygenase-like enzyme encoded by *hemO*. The presence and exchange of *hmbR* with gene cassettes probably encoding other iron-binding proteins in pathogenic *Neisseria* species may also underline the extensive recombination activity of this species.

**Other Gram-Negative Pathogens**

LEE of *Citrobacter rodentium*. LEE has also been identified in *C. rodentium*, the causative agent of transmissible murine colonic hyperplasia in suckling mice. Infected with *C. rodentium* develop mild diarrhea and coat ruff with an extensive recombination activity of this species.

Other bacterial species, such as iron, in the human environment. Pathogenic *Neisseria* species express, instead of siderophore systems, at least four outer membrane protein receptors that are specific for transferring (*bhpAB*), lactoferrin (*lbpAB*), hemoglobin (*hmbR*), and Hb-Hp-complexes (*hpuAB*). After being translocated into the cytoplasm, iron is released by a heme oxygenase-like enzyme encoded by *hemO*. The presence and exchange of *hmbR* with gene cassettes probably encoding other iron-binding proteins in pathogenic *Neisseria* species may also underline the extensive recombination activity of this species.

### PAI of *Bacteroides fragilis*

*B. fragilis* is a colonic commensal in the majority of adults and has been identified as the leading anaerobic isolate from clinical specimens, bloodstream infections, and abdominal abscesses (255, 275, 289). Its prevalence in anaerobic infections has been attributed mainly to its colonization of the gut (50%), suggesting a low level of recombination.

Particular strains of *B. fragilis*, designated enterotoxigenic *Bacteroides fragilis* (ETBF), have been associated with diarrhea in livestock, young children, and adults, and have also been isolated from patients with active inflammatory bowel disease (247, 278, 298). The primary pathogenicity factor of ETBF is the *Bacteroides fragilis* toxin (BFT, also known as fragilysin), a 20-kDa zinc-dependent metalloprotease belonging to the intramolecular chaperone protease family (165). BFT alters the morphology of certain human intestinal carcinoma cells, e.g., HT29C1, and causes fluid accumulation in ligated intestinal loops of lambs, rats, rabbits, and calves (248, 315). The target of BFT is the eukaryotic cell surface protein E-cadherin (375). In a current model of BFT-mediated disease, adherent ETBF strains secrete BFT, which diffuses through the tight junctions (zonula occludens) of intestinal cells to reach its target substrate, E-cadherin (zonula adherens). The extracellular domain of E-cadherin is then cleaved, resulting in cellular activation and cleavage of the extracellular E-cadherin domain, rearrangement of F-actin, and nuclear signaling with production of the inflammatory cytokine IL-8. The F-actin rearrangement results in diminished epithelial barrier function and activation of apical membrane ion transporters yielding net intestinal secretion. Moreover, secretion of IL-8 stimulates a submucosal inflammatory response that may also contribute to the net intestinal secretion (314). Three different BFT have been characterized from a piglet isolate, a human isolate, and a lamb isolate (for a review, see reference 314).

The *bft* gene is located within a 6,033-bp DNA region that is not present in non-toxigenic *B. fragilis* strains (NTBF) strains. This region contains virtually perfect 12 bp-direct repeats near its end and carries, in addition to *bft*, a gene for another metalloprotease, the deduced amino acid sequence of which is 28.5% identical and 56.3% similar to BFT and contains similar sequence motifs (242). It was termed metalloprotease II (MP II). This DNA region has been termed the *Bacteroides fragilis* pathogenicity island (BfPAI). Analysis of the nucleotide sequence of the left side of BfPAI revealed a DNA region containing the two mobilization genes *bfmA* and *bfmB*, which are organized in operons, and a putative origin of transfer (oriT). This region has the same genetic organization as the mobilization region of the 5-nitromimidazole resistance plasmid pIP417 and the clindamycin resistance plasmid pBFTM10 (99). The right end of the PAI contains the *bfmC* gene, whose product shows similarity to TraD mobilization proteins encoded by F- and R100 *E. coli* plasmids. Nucleotide analysis of that region in NTBF strains revealed that *bfmB* and *bfmC* are separated by a 16-bp GC-rich repeat. This sequence is precisely replaced by BfBAP in ETBF strains, so that the island is located between the stop codons of these genes. The G+C content of BfPAI (55%) and the flanking DNA (47 to 50%) differs from that of the *B. fragilis* chromosome (50%), suggesting that these are different genetic elements originating from different organisms (99). Maximal expression of BFT requires the presence of both the 6-kb BfPAI and its flanking DNA (100).

### rag locus of Porphyromonas gingivalis

*P. gingivalis* is a gram-negative anaerobic rod and is widely considered to be major pathogen in the development of destructive periodontal diseases such as chronic or aggressive periodontitis. *P. gingivalis* is able to express a number of pathogenicity factors including adhesins, LPS, capsule, collagenases, and other proteases (199, 202). It adheres to epithelial cells, red blood cells, collagen complexes, and other bacteria and, furthermore, has been found to invade host cells, e.g., epithelial cells (62). During a case-control study of the specific immunoglobulin G antibody response to *P. gingivalis* surface components, reaction of human sera with a bacterial 55-kDa protein led to the production of a monoclonal antibody and subsequently to the discovery of a genetic locus of *P. gingivalis* strain W 50 encoding this anti-
gen. The gene encoding the 55-kDa antigen was detected on a 9.5-kb Sau3AI DNA fragment and was designated ragB (1,506 bp) for “receptor antigen B.” It is located immediately downstream of ragA (3,054 bp), a gene whose product has homology to a class of TonB-dependent outer membrane receptors. This locus has a G+C content of 41%, in contrast to the typical G+C content of 47 to 48%. The rag genes are flanked by IS126 on one side and an incomplete ORF3, homologous to a V. cholerae transposase gene, on the other side (133). The rag-encoded proteins are probably lipoproteins in the outer membrane of P. gingivalis that act in concert at the surface of the bacterium to facilitate active transport of as yet unknown ligands (61). It could be shown that P. gingivalis strains containing the rag locus were isolated frequently from sites with increased pocket depth (>5 mm) but were absent from sites with depths of <4 mm and is therefore suggested to be a novel PAI.

Hrp island of Pseudomonas syringae. The plant pathogen P. syringae is able to grow to high population levels in leaf intercellular spaces of host plants and then to produce necrotic lesions. In nonhost plants or plants that express a race-specific resistance, the bacteria cause a hypersensitive response, a rapid, defense-associated programmed cell death of plant cells in contact with this pathogen (3). These effector functions are directed by the hrc (for “hypersensitivity and pathogenicity”) and hrc (for “hypersensitivity and conserved”) genes, which encode a T3SS, and by the avr (virulence) and hop (hpr-dependent outer proteins) genes, which encode effector proteins injected into plant cells. These effectors may also make the pathogen visible for the HR-triggering R-gene surveillance system of the host (179). The hrc and hrc genes are probably universal among necrosis-causing gram-negative plant pathogens, and they have been characterized in P. syringae, Erwinia amylovora, Xanthomonas campestris pv. vesicatoria, and Ralstonia solanacearum (3, 57). It was shown that the hpr and hrc genes are part of a PAI that has been designated Hrp (2). Hrp comprises the tree loci EEL (for “exchangeable effector locus”), the hpr-hrc gene cluster, and CEL (for “conserved effector locus”). Hrp is approximately 21 kb long, has a mosaic structure, and is integrated downstream of tRNA<sup>Leu</sup>.

Vap loci of Dichelobacter nodosus. The gram-negative anaerobe D. nodosus is the causative agent of foot rot in sheep. Strains of high, intermediate, and low virulence were identified, and virulence characteristics correlated with the presence of a chromosomal region containing the vap (virulence-associated protein) genes (178). Further analyses of the vap locus identified an insertion of 11.9 kb (52). The vap locus is inserted at a tRNA<sup>Asp</sup> locus and flanked by the intA gene encoding an integrase. Repetitive sequences up to 103 bp were located within the PAI. The presence of a further int gene indicated that loci is likely to be the result of two independent insertions that were distinguished as vap region 1 and vap region 3 (52). Hybridization analyses indicate the presence of a second intA gene in virulent D. nodosus strains. This second intA gene belongs to smaller chromosomal insertion at another tRNA<sup>Sec</sup> locus, harbors further vap genes and toxA, and was termed vap region 2 (24, 370). The association of vap regions with int genes might indicate the transfer of these regions by bacteriophages (51). Due to the lack of a genetic system for D. nodosus, the role of the vap PAI in virulence could not be confirmed experimentally.

PAI of Gram-Positive Pathogens

Listeria monocytogenes

Only two of the six Listeria species currently recognized cause listeriosis, L. monocytogenes and L. ivanovii. Whereas L. monocytogenes can infect humans and a wide range of animals including mammals and birds, L. ivanovii is mainly pathogenic for ruminants. Listeriosis is a food-borne infection, whose serious complications include sepsis and meningitis. It occurs mainly in immunocompromised patients, such as transplant recipients, infants and elderly persons receiving chemotherapy, individuals with diabetes or liver disease, and patients with human immunodeficiency virus infections.

About 25% of invasive listeriosis occurs in pregnant women. Maternal listeriosis results in fever, influenza-like syndromes, back pain, and headache. Spontaneous abortion, stillbirth, death of newborns within hours after birth, and neonatal sepsis are frequent results of this maternal infection (356, 371). Some infections occur without symptoms.

After ingestion of Listeria-contaminated food, bacteria pass through the stomach and cross the intestinal barrier, presumably via M-cells. They are then transported by the lymph or blood to the mesenterial lymph nodes, the spleen and the liver. L. monocytogenes and L. ivanovii are facultative intracellular pathogens, which are able to replicate in macrophages and a variety of nonphagocytic cells such as epithelial and endothelial cells and hepatocytes. After entry into the cell, Listeria escapes early from the phagocytic vacuole, multiplies in the host cell cytosol, and then moves through the cell by induction of actin polymerization. The bacteria then protrude into cytoplasmic evaginations, and these pseudopod-like structures are phagocytosed by the neighboring cells (356, 371).

LIP-1. The key steps in the intracellular life cycle of L. monocytogenes are basically determined by six genes, which are physically linked in a chromosomal island with of 9 kb (re-viewed in reference 354). This region consists of three transcriptional units. The middle portion carries the hly gene, which encodes the pore-forming listeriolysin O (LLO), a thiol-activated hemolysin, which is able to lyse erythrocytes and other cells in a cholesterol-dependent manner (for a review, see reference 356). LLO is an essential virulence factor of L. monocytogenes (356), and its inactivation leads to avirulence. The action of LLO is needed to disrupt the phagocytic vacuole for release of the bacteria into the cytoplasm. A current model for the escape of L. monocytogenes from the vacuole propose two functions for LLO: one is to dissipate the pH gradient and hence to halt maturation of the vacuole, and the other is to form a channel for the transfer of proteins, e.g., the phospholipases, or host proteins from the vacuole and act on the vacuole, leading to its dissolution. In addition, host proteins such as proteases may be involved in this process (223). There is evidence that LLO is also a multifunctional virulence factor with important roles in host-parasite interactions. In vivo and in vitro studies have shown that it triggers a number of host cell responses such as cell proliferation and focus formation in transfected fibroblasts, activation of the mitogen-activated pro-
tein kinase pathway in epithelial cells, mucus exocytosis induction in intestinal cells, modulation of internalization via calcium signaling, cytokine expression in macrophages, degranulation and leukotriene formation in neutrophils, apoptosis in dendritic cells, NF-kB activation, and expression of cell adhesion molecules in infected endothelial cells (356).

The genes mpl, actA, and plcB are located downstream from hly and are organized in an operon structure. ActA is a surface protein, necessary for actin-based motility and for spread from cell to cell. The ActA protein can be arbitrarily divided into three domains: an N-terminal domain rich in cationic residues, a central region with proline-rich repeats, and a C-terminal domain with a highly hydrophobic region that anchors the molecule in the bacterial surface (69, 193, 356). ActA acts as a scaffold to assemble the actin polymerization machinery of the host cell (41). The N terminus of ActA binds to monomeric actin and acts as a constitutively active nucleation-promoting factor by stimulating the intrinsic actin nucleation activity of the Arp2/3 complex. Also involved in this process are members of the eukaryotic enabled/vasodilator-stimulated phosphoprotein (Ena/VASP) protein family (205, 328). In a current model of L. monocytogenes actin-based motility, the eukaryotic Arp2- Arp3 complex is activated by the aminoterminus of ActA. ActA, Mena/VASP proteins, and profilin may be involved in Arp2-Arp3-mediated nucleation by delivering monomers to this complex. In this state, actin clustering can be observed around the bacterium. Once an actin filament is nucleated, profilin may increase nucleation rates. The Arp2-Arp3 complex caps the pointed ends of filaments and competes with capping protein. Mena/VASP proteins remain concentrated at the bacterium-tail interface, whereas capping protein can be found throughout the tail. Rapidly growing barbed ends are concentrated at the site necessary for force generation, a prerequisite for movement of bacteria through the cytosol (277). ActA is also essential for Listeria pathogenicity.

PlcB is a zinc-dependent broad-substrate-range phospholipase C, which is similar to the clostridial phospholipase C (alpha-toxin). This enzyme is secreted in an inactive form and then extracellularly modified by the Mpl protease, which is similar to PlcB, a zinc metalloenzyme. PlcB is, together with Hly, involved in the disruption of the primary vacuole formed after phagocytosis of extracellular Listeria cells (222, 327, 355).

The main function of PlcB is, however, the dissolution of the double-membrane secondary phagosomes formed after cell-to-cell spread. All three proteins are involved in a single process: the cell-to-cell spread. By this function, the pathogens can avoid the extracellular environment and can escape humoral effectors of the immune system during their dissemination in the host (276).

The plcA-prfA operon is located upstream of hly and is transcribed in the opposite direction. It encodes PlcA, a phosphatidylinositol-specific phospholipase C, and PrfA, a transcription factor. Whereas PlcA also has a destabilizing function on the membrane of the primary phagosome, PrfA transcriptionally regulates the expression of all genes encoded in this gene cluster, including its own expression.

The hly virulence gene cluster described above is stably inserted at the same position in the chromosome of L. monocytogenes and L. ivanovii but is not present in L. innocua and L. welshimeri. It has a DNA composition similar to that of the rest of the Listeria chromosome, lacks obvious traces of mobile genetic elements, and is not integrated adjacent to a tRNA gene. The PAI ends also do not possess direct repeats. Only two small ORF at one end of the island with low homology to phage ORF indicate an ancient mobility and indicate that the 9-kb virulence gene cluster is an ancient PAI that had lost its mobility. The authors named this island LIPI-1 for “Listeria pathogenicity island 1” (356).

LIPI-2. Beside LIPI-1, a second island of 22 kb, termed LIPI-2, has been described. LIPI-2 contains a cluster of several small, secreted internalin genes and the smIC gene, encoding the L. ivanovii sphingomyelinase C. LIPI-2 is specific for L. ivanovii and may play a role in the tropism of this pathogen for ruminants (117).

Internalin islets. Internalins comprise a family of well-characterized Listeria proteins, involved in the internalization of Listeria by susceptible cells that are normally not phagocytic (31, 107, 208). Internalins consists of a characteristic domain containing a variable number of leucin-rich-repeat regions constituting a right-handed helix (parallel b-helix). Two internalin subfamilies are already known. The first group consists of large proteins (70 to 80 kDa) that are attached via the C-terminal region to the cell wall. This group is represented by InlAB and by at least six other members (InIC2, InID, InIE, InIF, InIG, and InIH) (74, 107, 281). The second group consists of smaller proteins (25 to 30 kDa). They lack a C-terminal anchor region and are released in the extracellular medium. The prototype is InIC (70, 86). The known gene loci encoding internalins contain two to several internalins and have been termed internalin islets (354). Although internalins have been demonstrated to be relevant for Listeria virulence, their role in the pathogenesis of listeriosis is not fully understood.

Other loci. Other virulence-associated loci encode the stress tolerance mediators of the Clp group involved in intraphagosomal survival, such as ClpC, an ATPase belonging to the HSP-100 family, ClpP, a stress protease, and ClpE, also belonging to the family of HSP-100 stress proteins (108, 249, 294). Furthermore, the Ami protein is involved in attachment to host cells and the PrfA-dependent HPT hexose phosphate transporter (238, 291) is necessary for efficient intracellular proliferation. Whether these factors are encoded on genomic islands is not yet clear.

The genomes of L. monocytogenes and L. innocua have already been sequenced (116), and comparison of these genomes with those of other Listeria species will presumably lead to the discovery of novel islands involved in the specific life-style of Listeria and the evolution of pathogenicity.

**Staphylococcus aureus**

*S. aureus* is a common commensal found on human skin and respiratory tract mucosal surfaces. However, it is also a pathogen that is associated with a tremendous variety of human diseases ranging from self-limiting skin infections to life-threatening pneumonia or sepsis; nosocomial infections by *S. aureus* are common. A variety of toxins, such as hemolysins, staphylococcal exotoxins (Set), and superantigens (enterotoxins, exfoliative toxin, toxic shock syndrome toxin), are major virulence factors of *S. aureus*. These toxins are involved in the pathogenesis of staphylococcal diseases such as food poison-
ing, scalded skin syndrome, and toxic shock syndrome. Treatment of *Staphylococcus* infections becomes increasingly difficult, since resistance to a growing number of antibiotics has been observed in clinical isolates.

Genetic analysis of the virulence and resistance genes of various clinical isolates of *S. aureus* revealed that these genes are clustered in prophages and transposons, as well as with genomic islands that have characteristic features of PAI (for a review, see reference 257). The availability of three genome sequences of methicillin-resistant *S. aureus* (MRSA) strains allowed a detailed comparison of the structure of PAI in *S. aureus*. A remarkable difference between PAI in *S. aureus* and PAI in gram-negative pathogens is the presence of large gene clusters with allelic variations of specific toxins, proteases, and enzymes involved in pathogenesis. The allelic forms may allow adaptation of the pathogen to the various host environments that are colonized during infection. Given the large number of different PAI in *S. aureus* and the variability in structure and gene content, it is likely that characterization of further clinical isolates will lead to the detection of an increasing number of PAI in this group of pathogens.

*Staphylococcus cassette chromosome mec*. A major problem in the treatment of *S. aureus* infections is the presence of resistance to multiple antibiotics. Meticillin is the first semisynthetic, β-lactamase-resistant penicillin for therapeutic use. Resistance to meticillin in MRSA strains is conferred by penicillin-binding proteins and is usually accompanied by resistance to variety of other β-lactam antibiotics. It was observed that meticillin resistance is encoded by the chromosomal gene *mecA*. Analysis of the genome of MRSA strains revealed that antibiotic resistance genes are located within an unstable locus that has certain features of a PAI. The genomic island *Staphylococcus* cassette chromosome mec (SCC mec) carries the meticillin resistance determinants *mecI*, *mecR*, and *mecA*. The number of resistance genes located in SCC mec can vary among different strains and is dependent on the presence of transposons within SCC mec. In MRSA strains, the insertion of Tn 554 into SCC mec confers resistance to spectinomycin and erythromycin. The availability of genome sequences of several *S. aureus* strains indicated that the size and gene composition of SCC mec are highly variable and led to the distinction of allelic forms of SCC mec into types I to IV (9); the size of SCC mec ranges from 20.9 to 66.9 kb depending on the allelic form (64). Type I and II SCC mec were identified in nosocomial isolates of MRSA.

In addition to resistance genes, SCC mec encode the recombinases CcrAB (177) and is flanked by att sites. Under experimental conditions, CcrAB can promote the excision of SCC mec from, or the site-specific integration into, the chromosome. Other genes present in the island are apparently not associated with virulence. The presence of SCC mec in a number of staphylococcal species has led to the suggestion that it is transmissible. Although *mecA* could be could be transferred experimentally by phage transduction to other staphylococcal species, no naturally occurring transducing bacteriophage which is capable of transferring genetic material across the species barrier has been described (177).

**PAI encoding toxic shock syndrome toxins.** In addition to SCC mec, a larger number of genomic islands have been identified. It was observed that chromosomal *tst* genes encoding toxic shock syndrome toxins (TSST) are located on mobile genetic elements that were distributed among *S. aureus* strains. TSST-1 and enterotoxins of *S. aureus* function as superantigens (10). Staphylococcal superantigens are a group of high-molecular-weight pyrogenic proteins that are potent stimulatory agents for CD4+ T lymphocytes. As such, they can have profound effects on the immune system. They stimulate T cells by cross-linking the variable part of the β-chain of the T-cell receptor with major histocompatibility complex class I molecules on accessory or target T cells outside the peptide-binding groove area (96). This will result in a nonspecific activation of a large proportion of T cells. Staphylococcal superantigens are associated with food poisoning, toxic shock syndrome, multiple sclerosis, Kawasaki’s disease, and atopic allergy (reviewed in reference 351).

Analysis of the flanking regions of the *tst* gene identified a PAI that has been termed SaPI1 (212). Beside *tst*, a gene encoding a homologue of VapE of *D. nodosus* and a putative enterotoxin have been identified. SaPI1 is 15,233 bp long and is inserted in an *attC* site close to the *tst* gene. Remarkable features of SaPI1 are its mobility and instability. The excision of SaPI1 from the chromosome and its presence as episomal DNA was observed. This study also identified a second locus carrying a *tst* gene that was referred to as SaPI2. Transduction of SaPI1 and SaPI2 by particles of helper phages was demonstrated (212, 296). In the absence of these helper phages, these islands remain stably integrated in the chromosome.

A PAI related to SaPI1 was identified in a bovine isolate of *S. aureus* (93). SaPIBov is 15,891 bp long and is inserted at the 3′ end of the GMP synthase gene (*gmps*). It harbors the enterotoxin genes *sec* and *sel* as well as the *tst* gene. A third PAI, termed SaPI3, was identified by Yarwood et al. (378). This locus contains *sek* and *seq*, encoding two novel enterotoxins, and the overall structure is similar to that of SaPI1. The *att* sites flanking the PAI are identical; however, a *tst* gene is absent from SaPI3.

**vSa families of PAI.** Comparative analyses of the genomes of six *S. aureus* isolates defined six families of PAI and led to the renaming of previously identified PAI to vSa1, vSa2, vSa3, vSa4, vSaα, and vSaβ (9), where v stands for “vaccine.” Depending on the authors, the loci have different designations; for consistency we follow the most comprehensive nomenclature suggested by Baba et al. (9). Islands vSa1 to vSa4 harbor integrate genes as putative elements of genetic mobility. Furthermore, these PAI are flanked by *att* sites. In contrast, members of the vSaα and vSaβ families harbor transposase genes, indicating that transposons may have been the origin of these elements.

1. **vSa1 to vSa4.** The designation vSa1 was suggested for a family of genomic islands including the vSa1 locus of community-acquired *S. aureus* MW2 and PAI previously termed SaPI1 and SaPI3 (Fig. 5D). PAI of this family harbor large clusters of genes for enterotoxins and encode TSST (see above). The vSa2 family comprises vSa2 of community-acquired MRSA and SaPIBov of a bovine *S. aureus* isolate. Like vSa1, these loci harbor genes encoding enterotoxins and TSST (described above). vSa3 has been identified in community-acquired MRSA strain MW2 and nosocomial MRSA strains Mu50. The type II vSa3 of strain MW2 harbors novel allelic forms of the enterotoxin genes *sel* and *sec*. The spontaneous excision of
vSa3 from the chromosome and formation of an episomal circular DNA occurs frequently. The vSa4 family includes several allelic forms of a genomic island in nosocomial and community-acquired MRSA infection. Loci previously referred to as the TSST-1 islands SaPln1 and SaPlm1 of nosocomial MRSA isolates are now referred to as type I vSa4. Type I vSa4 of nosocomial isolates harbors virulence genes encoding allelic forms of enterotoxin sel and sec and the tst gene encoding TSST. In contrast, the tst gene is absent in type II vSa4 of community-acquired MRSA. Spontaneous excision of vSa4 was observed, but the frequency was lower than that of excision of vSa3 (9).

(ii) vSaα. The vSaα islands are present in all sequenced S. aureus genomes, but the size and number of ORF in this group is variable. Exotoxin islands previously designated SaPln2 and vSaPlm2 (200) are members of the vSaα family. The loci contain up to 11 allelic forms of set genes for putative staphylococcal exotoxins, the lukDE genes encoding leukotoxins, and lipoprotein gene clusters.

(iii) vSaβ. Like vSaα, the vSaβ genomic islands are present in all sequenced genomes, but the size and gene composition of the various vSaβ islands are highly variable among isolates. Enterotoxin islands previously referred to as SaPln3 and SaPlm3 (200) belong to this family. A cluster of genes encoding serine proteins is present in all varieties of vSaβ. Superantigen gene clusters are observed in nosocomial MRSA strains but absent in community-acquired MRSA strains. The bsa gene encoding a putative bacteriocin is present only in type II vSaβ of community-acquired MRSA strains, and this toxin could be important for competition with other bacteria in an environment outside the host.

Neither vSaα nor vSaβ is spontaneously excised from the chromosome. In these islands, transposase genes are present that are inactive due to frameshifts.

There are also several prophages in the genomes of various S. aureus isolates that harbor various genes encoding toxins. The presence of int genes and att sites in the PAI of S. aureus suggests a close relationship between PAI and bacteriophages in S. aureus, indicating that several of the PAI in this species have been acquired in form of phage genomes. Baba et al. (9) classified prophages δSa1, δSa2, and δSa3. Novel allelic forms of the luk genes for leukotoxins of the “Panton-Valentine leukocidin” class are present on δSa2. The enterotoxin genes sea, seg, sek, and sak, encoding staphylokinase, have been identified in members of the δSa3 family.

edt PAI. The analysis of loci of genes encoding exfoliative toxins ETD and EDIN-B recently led to the identification of a further locus termed edt PAI (377). The edt PAI in clinical S. aureus isolate TY114 is 9,054 bp long and absent in MRSA strain N315. This locus contains edt and edin-B, encoding the exfoliative toxins. The purified EDT toxin exhibited exfoliative activity in a murine model (377). The locus also contains a restriction/modification system and an IS element and is flanked by short DRs. A linkage to edt was observed in all clinical isolates positive for edin-B, and the edt PAI is also frequently distributed among strains that are not associated with scalded skin syndrome or other exfoliative forms of S. aureus infections.

Streptococcus spp.

As with staphylococci, a variety of diseases are linked to infections with streptococci. Diseases caused by group A streptococci (S. pyogenes) include skin infections, pharyngitis, and the highly progressive necrotizing fasciitis. Group B streptococci (S. agalactiae) are frequently associated with neonatal septic infections. Pneumococci (S. pneumoniae) are frequent commensals on the mucosa of the upper respiratory tract, but they can also cause pneumonia, otitis media, and other infections of the respiratory tract.

In contrast to S. aureus, PAI appear less relevant to the evolution of virulence in streptococci. The genome sequence of S. pyogenes revealed that a large number of virulence factors have been acquired by bacteriophage-mediated horizontal gene transfer but that PAI are absent from the genome (90). In contrast, a chromosomal region reminiscent of a PAI has been identified in S. pneumoniae. This locus was termed pneumococcal pathogenicity island 1 (PPI-1) (34). PPI-1 is about 27 kb long, is absent in other streptococci and contains transposable and recombine genes that may be related to the mobility of the locus. A virulence factor identified in PPI-1 is an iron uptake system encoded by pit2ABCD. There is also the related pit1 operon outside of the PAI, but full virulence of pneumococci in pulmonary and systemic animal models requires the function of both the Pit1 and Pit2 iron uptake systems (34).

Enterococcus faecalis

E. faecalis is a normal commensal in the human intestinal flora but can also lead to nosocomial infections. Multiple antibiotic resistance is frequently observed in E. faecalis, and vancomycin resistance was first reported for E. faecalis isolates in an outbreak in a hospital (reviewed in reference 325). Known virulence factors of E. faecalis include the toxin cytolysin (Cyl) and the surface protein Esp. Esp function is important for biofilm formation, and the presence of Esp is linked to high virulence of clinical isolates (322). The loss of both cyl and esp has been observed, and the deletions were located within the same chromosomal region. Detailed sequence analysis of this region revealed the presence of a 154-kb PAI that contains the cyl operon, esp, and several genes of unknown function (321). The locus has a G+C content of 32.2%, lower than that of the core genome of E. faecalis, and shows an intrinsic genetic instability. Genes related to the transfer and mobility of the PAI include a transposase and a traG-like gene. About one-third of the PAI shows sequence similarity to conjugation-related related genes of enterococcal plasmids, indicating that part of the PAI could have originated from chromosomal integration of a plasmid. Deletions of parts of the PAI were observed, but transfer of the PAI by conjugation could not be observed under experimental conditions.

PaLoc of Clostridium difficile

The genus Clostridium comprises strictly anaerobic, endospore-forming rods, including several species that are important human pathogens. C. difficile is frequently found in the physiological intestinal flora of children younger than 1 year and is occasionally isolated from healthy adults. However, it is also
involved in antibiotic-associated diarrhea and pseudomembranous colitis, a sometimes fatal syndrome. The administration of antibiotics can induce the production of potent toxins by *C. difficile*, which are the major virulence factors. Toxin A is an enterotoxin that induces fluid secretion into the intestinal lumen. Toxin B is a highly active cytotoxin that causes severe damage to epithelial cells of the intestinal mucosa. The massive death of intestinal cells accompanied by infiltration of inflammatory cells results in the formation of a pseudomembranous layer of dead cells that compromises the normal resorption of the colonic mucosa.

The comparison of toxigenic and nontoxigenic strains of *C. difficile* revealed the presence a distinct locus that contains the toxin genes *tcdA* and *tcdB* (32). This locus was termed PaLoc (for “pathogenicity locus”), consists of five ORF, and forms an insertion of 19.6 kb between genes that are contiguous in nontoxigenic *C. difficile* strains. There are no genes associated with genetic mobility or instability located within PaLoc; and the genetic mechanisms that resulted in the presence of toxigenic and nontoxigenic strains have not been identified so far.

The formation of toxins by *C. difficile* is strictly associated with the presence of PaLoc (32, 55). Comparison of highly virulent isolates of *C. difficile* to other toxigenic isolates indicated that insertions and deletions in PaLoc can result in a loss of toxin A expression and altered levels of toxin B expression (331). Analysis of the expression of PaLoc genes indicate a growth phase-dependent expression of toxins in the stationary phase and a function of PaLoc-encoded TcdC and TcdD as negative and positive regulators of PaLoc gene expression, respectively (163, 334). The genome-sequencing projects of *C. perfringens* (326) and *C. tetani* (35) did not indicate the presence of PaLoc.

### CONCLUDING REMARKS

#### Approaches to the Identification of New PAI

The majority of PAI described in this review have been identified by the characterization of the genetic organization of previously known virulence factors. However, the availability of genome data provides a basis for directed approaches to the identification of PAI. Sequencing of entire bacterial genomes and whole-genome sequence analysis and comparison have permitted a deeper insight in the structure and properties of bacterial chromosomes at a revolutionary speed. Deployment of computer software supports the discovery of novel genomic islands in many bacteria. The G+C content of different parts of the genome (such as protein-encoding genes, RNA genes, spacers, promoters, regulatory regions, genomic islands, and prophages) reveals different positive linear correlations with the G+C content of the entire genomic DNA (245). Plots of G+C contents are extensively used in comparative analysis of complete genomes. Other typical analyses at the genome sequence level include the GC skew, i.e., (G+C)/(G−C), the assessment of dinucleotide and tetranucleotide relative abundance values (δ-difference), the identification of rare and frequent oligonucleotides, and the evaluation of codon usage biases (121, 176). Other interesting methods are the continuous wavelet transform to analyze long-range correlations associated with G+C patterns in DNA sequences (6) or the generation of a wavelet scalogram (213). Since most islands have a different G+C content from the core genome, they can be discovered by such mathematical models.

Another option is the scanning of potential insertion sites of foreign elements. Since many PAI are inserted at tRNA genes, it can be of interest to examine the vicinity of tRNA genes for the presence of pathogen-specific insertions. This approach was used for the identification of genomic regions specific to *S. enterica* serovars Typhimurium and Typhi (135). When genome sequences are not available, it is still possible to clone and identify pathogen-specific genomic elements. The principle of DNA hybridization can also be used for this purpose. Genomic regions present in both species will be able to form hybrids between DNA strands of both species, while elements absent in one species can hybridize only to the complementary strand. These fragments can be directly cloned or used for amplification of the species-specific DNA by PCR. Another method is based on the instability of PAI. A marker for negative selection is integrated into the bacterial chromosome by a random-mutagenesis approach. If selective pressure is applied, the only mutants that will survive are those that have eliminated the selection marker or the region containing the marker. This approach thus allows the identification of rather rare events of deletion of a PAI. The technique was applied to *S. flexneri* and resulted in the identification of SH1 (initially defined as *she* PAI [282]). In addition to identification of new PAI, this approach is useful in quantifying the genetic instability of known PAI.

It should be mentioned that the genomic approaches are usually not suitable to distinguish between horizontally acquired virulence genes and genes with other functions. These techniques can lead to the identification of genomic islands; however, further experimental analyses are usually required to investigate the contribution of these elements to virulence. Other approaches have been devised to identify virulence genes based on their function (148) or expression in vivo (219), but these approaches cannot discriminate between PAI and loci that are present in the core genome of the pathogen.

### Evolution of PAI

Our description of PAI in various bacterial pathogens demonstrates the variability in the structure and evolution of this class of genetic elements. There are examples of highly mobile PAI that can be converted into phages or plasmids. On the other hand, we have described PAI that have stably integrated into the chromosome and already lost many characteristics of PAI. These elements turned into components of the core genome. Between these extremes, various intermediate forms of PAI can be found.

Some PAI still contain bacteriophages or transposon genes, indicating the mode of horizontal gene transfer during their acquisition. We have described several examples that highlight the role of bacteriophages in the transfer of virulence determinants during bacterial evolution. One extreme is probably VPI of *V. cholerae*, a PAI that can be converted into a lytic bacteriophage. Other PAI lack such characteristic mobility genes but harbor virulence genes that can be found in homologous form on virulence plasmids in other pathogens. For example, gene clusters encoding T3SS or pili can be found...
within PAI as well as on plasmids, indicating that the corre-
sponding PAI represent chromosomal integrations of plas-
mids. Obviously, there is a fluid transition between PAI and
bacteriophages, transposons, or plasmids.

Interestingly, many pathogens also harbor virulence loci that
were clearly acquired by horizontal gene transfer but do not
fulfill the criteria of PAI due to small size and/or lack of
generic instability or mobility. These elements were termed
pathogenicity islets by analogy to PAI and have made similar
contributions to the evolution of bacterial virulence. A remark-
able example is found in *S. enterica*, in which the functions of
the PAI SPI-1 and SPI-2 are complemented by clusters of small
islets (Fig. 6). These islets encode effector proteins for the
T3SS encoded by SPI-1 and SPI-2. While both SPI are stably
integrated into the chromosome, genes for effector proteins
are located in the genomes of functional or cryptic bacterio-
phages. This example demonstrates that a basic virulence func-
tion can evolve as a PAI and additional virulence determinants
subsequently are added by bacteriophages. These bacterio-
phages might form a mobile pool for virulence determinants. It
will be of interest to investigate if similar combinations of fixed
and mobile virulence genes are present in other pathogens.

### Bacterial Pathogens without PAI?

Although most groups of gram-negative and gram-positive
pathogens have been described in this review, the reader has
probably noticed the absence of several important groups of
pathogens. To our current knowledge, PAI are absent in *My-
cobacterium* spp., *Chlamydia* spp., the spirochetes, most strep-
tococcal species, and several other pathogens.

The reasons for the absence of PAI in certain species have
not been understood, but comparison of the life-style of patho-
gens with and without PAI might give some hints about the
underlying principles. Some groups of pathogens lacking PAI
show an extreme adaptation to a specific host environment that
is also accompanied by reduction of the genome size and loss
of the ability to replicate outside a host. In contrast, most
pathogens harboring PAI show a high degree of flexibility in
the utilization of different hosts or body sites of a host for their
proliferation. In addition, pathogens containing PAI are often
able to live in natural environments. These observations sug-
gest that PAI extend the spectrum of habitats that can be
colonized by a bacterial species. The observation that meta-
bolic functions can also be found in genomic islands supports
the notion that acquisition and maintenance of these elements
allows the access to new habitats. Consequently, highly
adapted and specialized pathogens evolve in the opposite di-
rection, i.e., toward loss of flexibility. The specialization is
accompanied by genome reduction, and it is possible that such a
reduction led to the deletion of major portions of horizontally
acquired DNA elements.

Extreme adaptation of a parasitic lifestyle might also result
in reduced access to the microbial gene pool. For example,
obligate intracellular pathogens have lost mechanisms that al-
low genetic exchange by conjugation or natural competence for
transformation. However, even highly adapted pathogens such as
*Chlamydia* spp. have specific bacteriophages (158) that
might contribute to a certain degree of horizontal gene trans-
fer.

We also find pathogens that lack PAI despite having a high
degree of flexibility in host range and a free-living life-style,
such as most streptococci. An explanation of this observation
might involve the high rate of recombination in this group of
bacteria. One can speculate that horizontal gene transfer also
plays an important role in these pathogens but that the rapid
recombination may make the identification of horizontally ac-
quired DNA difficult. The absence of PAI in most streptococ-
cal species could be explained by the high rate of genome
rearrangements in these bacteria, resulting in mosaic chromo-
somes and mosaic gene structures (132). The presence of mo-
bile genetic elements in streptococci indicates that the inser-
tion of large blocks of virulence genes may also take place. It
has been estimated that 10% of the total genome of *S. pyogenes*
consists of bacteriophages and transposons (90). The rapid
genomic rearrangements in streptococci may result in separa-
tion of horizontally acquired gene blocks, counteracting the
emergence of typical PAI.

In conclusion, uptake of large fragments of DNA containing
virulence genes is likely to occur in the majority of bacterial
pathogens. The mode of recombination of foreign DNA into
the host chromosome may determine whether or not a detect-
able PAI is formed.

### OUTLOOK

This review describes the features and role in pathogenesis
of PAI of important human pathogens. This description can
only be a snapshot (as of February 2003), and we are aware
that the number of PAI is constantly increasing. Moreover, not
all data required have been available for all PAI reviewed here,
so that there remain some gaps in our review, which will hope-
fully be filled in soon. The availability of a large number of
genome rearrangements of pathogenic bacteria and their benign
relatives currently offers a unique opportunity for the identifi-
cation of novel pathogenic specific genomic islands. However,
this knowledge has to be complemented by improved model
systems for the analysis of virulence functions of bacterial
pathogens.

In this review, we have described PAI that have been ac-
quired during the speciation of pathogens from their non-
pathogenic or environmental ancestors. However, the acquisi-
tion of PAI is not only an ancient evolutionary event that led
to the appearance of bacterial pathogens on a timescale of
millions of years. We have listed several examples in which the
acquisition of PAI is a mechanism that contributes to the
appearance of new pathogens within the life span of a human
generation. The most remarkable example is the emergence of
toxigenic strains of *V. cholerae* and the onset of recent cholera
pandemics.

In addition to specific virulence factors that interact with
host cell functions, we have described the phenomenon of
resistance islands, chromosomal insertions that encode one or
several antibiotic resistances. A prominent example is the SCC
*mec* locus of MRSA strains, and there are indications that
chromosomal islands are also involved in the emergence of
new multidrug-resistant strains of clinically important patho-
gens such as *S. enterica* and *Shigella* spp.

The acquisition of knowledge about PAI, their structure,
their mobility, and the pathogenicity factors encoded by PAI

not only is helping in gaining a better understanding of bacterial evolution and pathogen interaction with eukaryotic host cells but also can have important practical implications. Genes located within PAI have been used as diagnostic markers for the identification of pathogens in clinical specimen and for the differentiation of pathogenic strains from closely related non-pathogenic relatives. Toxins encoded by PAI genes can be used as tools in cell biology (304). Protein secretion systems can be used to deliver heterologous antigens for vaccination strategies with live carrier strains (113, 295). Virulence determinants used to deliver heterologous antigens for vaccination strategies as targets for novel forms of therapeutic intervention of bacterial infections.

The availability of microbial genome sequences has provided an extremely useful platform for future work. When this is combined with the understanding of the basic concepts of evolution of bacterial virulence and the crucial role of PAI in this process, microbiologists will be able to identify virulence traits of new emerging pathogens or strains very efficiently.

ACKNOWLEDGMENTS

Work in our laboratories was supported by grants (SCHM1360/12-1 and SCHM1360/1-3 to H.S and HE 1964/4-2 and HE1964/4-3 to M.H.) of the Deutsche Forschungsgemeinschaft priority program Ecology of Bacterial Pathogens.

REFERENCES


24. Buchkremer, F., B. Goujon, S. Röösli, and K. H. Meyer. 1998. The high-pathogenicity island of Yersinia pseudotuberculosis can be inserted into...


AUTHOR’S CORRECTION

Pathogenicity Islands in Bacterial Pathogenesis

Herbert Schmidt and Michael Hensel

Institut für Medizinische Mikrobiologie und Hygiene, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Dresden, and Institut für Klinische Mikrobiologie, Immunologie und Hygiene, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Volume 17, no. 1, p. 14–56, 2004. Page 40: Some of the text in the section entitled “Pathogenic Neisseria spp.” was used essentially verbatim from an article by J. P. Dillard and H. S. Seifert (Mol. Microbiol. 41:263–277, 2001) without appropriate attribution and without permission from the copyright owner. Dr. Schmidt assumes responsibility for this unintended omission and apologizes to the authors of the above-mentioned article and to the publisher of Molecular Microbiology.