Molecular Microbiology (1997) **23**(6), 1089–1097

MicroReview

Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution

J. Hacker,[1]* G. Blum-Oehler,[1] I. Mühldorfer[1] and H. Tschäpe[2]

[1]Institut für Molekulare Infektionsbiologie, Röntgenring 11, 97070 Würzburg, Germany.

[2]Robert-Koch-Institut, Bereich Wernigerode, Burgstraße 37, 38855 Wernigerode, Germany.

Summary

Virulence genes of pathogenic bacteria, which code for toxins, adhesins, invasins or other virulence factors, may be located on transmissible genetic elements such as transposons, plasmids or bacteriophages. In addition, such genes may be part of particular regions on the bacterial chromosome, termed 'pathogenicity islands' (Pais). Pathogenicity islands are found in Gram-negative as well as in Gram-positive bacteria. They are present in the genome of pathogenic strains of a given species but absent or only rarely present in those of non-pathogenic variants of the same or related species. They comprise large DNA regions (up to 200 kb of DNA) and often carry more than one virulence gene, the G+C contents of which often differ from those of the remaining bacterial genome. In most cases, Pais are flanked by specific DNA sequences, such as direct repeats or insertion sequence (IS) elements. In addition, Pais of certain bacteria (e.g. uropathogenic *Escherichia coli*, *Yersinia* spp., *Helicobacter pylori*) have the tendency to delete with high frequencies or may undergo duplications and amplifications. Pais are often associated with tRNA loci, which may represent target sites for the chromosomal integration of these elements. Bacteriophage attachment sites and cryptic genes on Pais, which are homologous to phage integrase genes, plasmid origins of replication or IS elements, indicate that these particular genetic elements were previously able to spread among bacterial populations by horizontal gene transfer, a process known to contribute to microbial evolution.

Received 8 June, 1996; revised 13 January, 1997; accepted 16 January, 1996. *For correspondence. E-mail j.hacker@rzbox.uni-wuerzburg.de; Tel. (931) 312575; Fax (931) 571954.

Introduction

Antibiotic-resistance factors, bacteriocins, specific metabolic functions, such as enzymes involved in the degradation of xenobiotic compounds, proteins involved in secretion, as well as bacterial pathogenicity factors can be encoded by mobile genetic elements. While the ST-enterotoxin genes of *Escherichia coli* are part of a transposon, genes encoding other toxins, such as Shiga-like toxins of *E. coli*, cholera toxin of *Vibrio cholerae*, diphtheria toxin of *Corynebacterium diphtheriae*, neurotoxins of *Clostridium botulinum* and cytotoxin of *Pseudomonas aeruginosa*, are phage encoded (Salyers and Whitt, 1994; Waldor and Mekalanos, 1996). Important virulence factors of Gram-negative pathogens (e.g. *Shigella flexneri*, *Salmonella* spp., *Yersinia* spp.) as well as of Gram-positive pathogens (*Clostridium tetani*) are plasmid encoded. These determinants have the capacity to spread among a certain population. Horizontal gene transfer by plasmids or phages may therefore play a role in the creation of new pathogenic variants. However, other virulence determinants are located on the chromosome, where they are often associated in so-called 'virulence blocks' or 'virulence cassettes'. Particular regions of chromosomally encoded virulence-associated genes have also been termed 'pathogenicity islands' (Pais) (Hacker *et al.*, 1990; Blum *et al.*, 1994). In this review we summarize data demonstrating that these islands represent unique genetic elements which contribute to bacterial virulence and to the genetic flexibility of certain bacterial species and thus may be involved in processes of microbial evolution.

Definition of pathogenicity islands

In the early 1980s, Goebel and coworkers analysed uropathogenic *E. coli* strains which were able to produce specific virulence factors, including α-haemolysin (Hly; Hacker *et al.*, 1983; Knapp *et al.*, 1985). It became obvious that the *hly* genes are located on large chromosomal DNA regions which were termed 'haemolysin islands' (Knapp *et al.*, 1986). Further studies revealed that additional genes involved in uropathogenicity are located on those islands (e.g. determinants coding for P fimbriae), which were consecutively renamed 'pathogenicity islands' (Hacker *et al.*, 1990; Blum *et al.*, 1994). As the term 'pathogenicity

island' (Pai) has become increasingly popular (Falkow, 1996; Lee, 1996; Groisman and Ochman, 1996), it is necessary to develop a precise definition. A Pai can be defined according to the following criteria.

(i) Carriage of (often many) virulence genes.
(ii) Presence in pathogenic strains, and absence or sporadic distribution in less-pathogenic strains of one species or a related species.
(iii) Different G+C content in comparison to DNA of host bacteria.
(iv) Occupation of large chromosomal regions (often >30 kb).
(v) Represent compact, distinct genetic units, often flanked by direct repeats.
(vi) Association with tRNA genes and/or insertion sequence (IS) elements at their boundaries.
(vii) Presence of (often cryptic) 'mobility' genes (IS elements, integrases, transposases, origins of plasmid replication).
(viii) Instability.

The question of whether the features of the definition fit the Pais of different species will be discussed in the following chapters.

Pathogenicity islands of *E. coli*

The first intensively studied Pais were those which are part of the genomes of uropathogenic and enteropathogenic *E. coli* strains (Hacker *et al.*, 1983; Low *et al.*, 1984; Mühldorfer and Hacker, 1994; McDaniels *et al.*, 1995; Blum *et al.*, 1994; Swenson *et al.*, 1996). The uropathogenic *E. coli* strains 536 (O6:K15) and J96 (O4:K6) carry Pais with sizes of 70–190 kb (see Table 1). The Pais of strain 536, indicated in Fig. 1, represent 'prototypes' for the pathogenicity islands discussed in this article. While Pai I encodes only α-haemolysin, Pai II carries the α-haemolysin (hly) determinant as well as the *prf* determinant, encoding 'P-related fimbriae'. The Pais may delete from the chromosome of the wild-type strains at frequencies of 10^{-4}, resulting in non-virulent mutant strains (Hacker *et al.*, 1990). Each of the two Pais of *E. coli* strain 536 are flanked by two direct repeats of 16 bp (Pai I) and 18 bp (Pai II), one copy of which remains in the chromosomes of the Pai-negative mutants (Knapp *et al.*, 1986; Blum *et al.*, 1994). These findings and the fact that the deletion of Pais of strain 536 is *recA* independent (A. Ritter, unpublished) argue for a contribution of a specific recombinase which facilitates illegitimate recombination, resulting in Pai-negative variants of strain 536. One copy each of the two direct repeats of the Pais are part of phage-attachment sites located at the 3′ end of tRNA genes. Thus, Pai I is flanked by the *selC* gene (82 units), encoding the selenocysteine-specific tRNA, while Pai II is associated with *leuX*, which codes for a minor leucyl $tRNA_5^{Leu}$ and is located at 97 units on the chromosome (Blum *et al.*, 1994). Interestingly, *leuX* acts as a regulator gene, influencing the expression of other virulence-associated genes, including those coding for type I fimbriae, enterochelin, serum resistance and flagella (Ritter *et al.*, 1995; Susa *et al.*, 1996). The fact that Pai II of the uropathogenic *E. coli* strain 536 is located next to *leuX*,

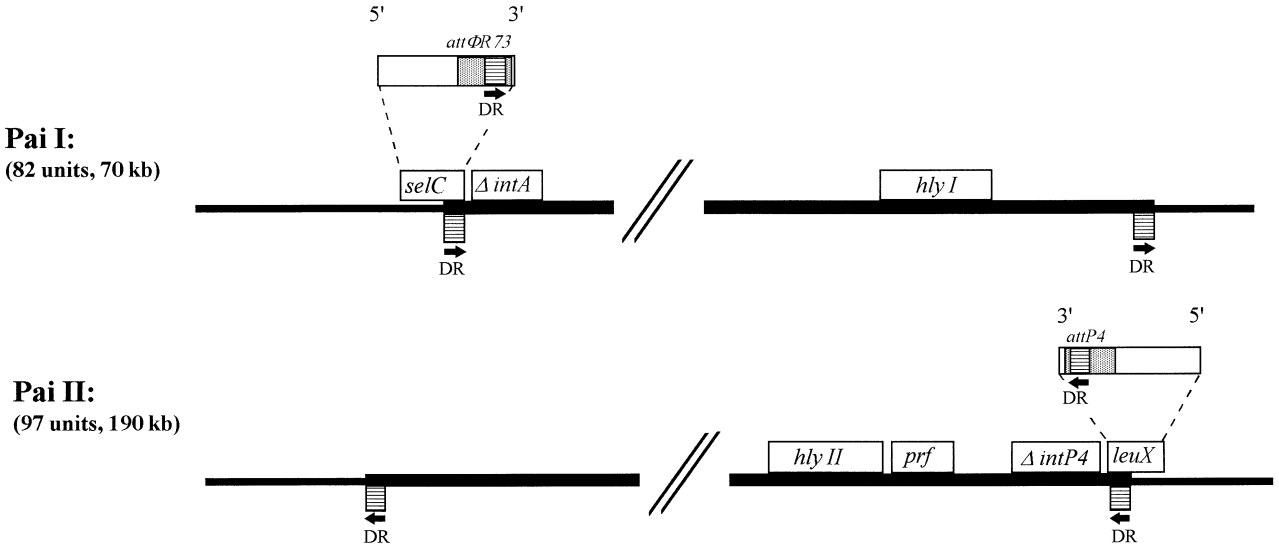

Fig. 1. Structures of pathogenicity islands (Pais) I and II of the uropathogenic *E. coli* strain 536. The black bars represent Pai-specific DNA. Direct repeats (DR) of 16 bp (Pai I) or 18 bp (Pai II) are indicated by horizontally striped boxes; their orientation is marked by arrows. Putative genes are indicated by open boxes. The exact structures of the tRNA genes are given above the *selC*- and *leuX*-specific boxes. Abbreviations: *selC*, selenocysteine-specific tRNA; Δ*intA*, cryptic integrase gene; *att* ΦR73, attachment site for the integration of the ΦR73 phage; *hly*, haemolysin gene cluster; *prf*, P-related fimbrial gene cluster; Δ*intP4*; cryptic integrase gene of P4 phage; *leuX*, leucine-specific tRNA; *attP4*, attachment site for the integration of the P4 phage.

Table 1. Pathogenicity islands (Pais) of Gram-negative bacteria.

Pathogenic microorganism	Pai designation	Relevant genes	G+C content[a]	Size (kb)	Location (units)	Boundary sequence	Associated tRNA	Deletion frequency	Reference
E. coli (536)									
Uropathogenic	Pai I	*hly*	51/41	70	82	DR 16 bp	*selC*	10^{-4}–10^{-5}	Blum *et al.* (1994)
Uropathogenic	Pai II	*hly prf*	51/41	190	97	DR 18 bp	*leuX*	10^{-4}–10^{-5}	Blum *et al.* (1994)
E. coli (J96)									
Uropathogenic	Pai IV	*hly pap*	51/41	170	64	–	*pheV*	?	Swenson *et al.* (1996)
Uropathogenic	Pai V	*hly prs cnfI*	51/41	110	94	DR 135 bp	*pheR*	10^{-5}	Blum *et al.* (1995); Swenson *et al.* (1996)
E. coli (E2348/69)									
Enteropathogenic	LEE (Pai III)	*eaeA espAB sepA–I*	51/39	35	82	–	*selC*	Stable	McDaniel *et al.* (1995)
Y. pestis	HPI (*pgm* locus)	*hms* HFRS *fyuA irpB–D*	46–50/46–50	102	–	IS*100*	–	10^{-4}–10^{-5}	Fetherston and Perry (1994)
Y. enterocolitica	HPI	*fyuA irp2 (irp1)*	46–50/56	45	–	–	*asnT*	10^{-5}	Carniel *et al.* (1996)
S. typhimurium	SPI I	*inv spa hil*	52/40–47	40	60.3	–	–	Stable/unstable[b]	Mills *et al.* (1995)
S. typhimurium	SPI II	Type III secretion genes, two-component genes	52/40–47	40	30.7	–	*valV*	Stable	Shea *et al.* (1996); M. Hensel & D. W. Holden, personal communication
H. pylori	Cag Pai	*cagA–T*	38–45/35	40	–	DR 31 bp	*glr*	Unstable	Censini *et al.* (1997)
D. nodosus	Vap region	*vapA–E*	45/52	11.9	–	DR 19 bp	*serV*	?	Cheetham *et al.* (1995)
D. nodosus	Vrl region	*vrl*	49/57	27	–	–	*ssrA*	?	Billington *et al.* (1996)
V. cholerae	TCP–ACF element	*acf tcp toxT int*	–	45	–	*att* sites	20 bp	Stable	Kovach *et al.* (1996)

a. G+C (%) content of the host organism versus G+C (%) content of Pai-specific genes.
b. Unstable in *S. senftenberg* and *S. lichtfield* (Galan and Sansonetti, 1996).
Abbreviations: *hly*, haemolysin gene cluster; DR, direct repeat; *selC*, selenocysteine-specific tRNA; *prf*, P-related fimbrial gene cluster; *leuX*, leucine-specific tRNA; *pap*, pili associated with pyelonephritis gene cluster; *pheV, pheR*, phenylalanine-specific tRNAs; *prs*, P-related sequence gene cluster; *cnfI*, cytotoxic-necrotizing factor I gene; LEE, locus of enterocyte effacing; *eaeA*, *E. coli* attaching and effacing gene A; *esp*, enteropathogenic *E. coli* secreted protein gene; *sep*, secretion of *E. coli* protein genes; HPI, high pathogenicity island; *pgm*, pigmentation locus; *hms* HFRS, haemin storage protein gene HFRS; *fyuA*, ferric yersiniabactin uptake A; *irp*, genes encoding iron-repressible proteins; IS, insertion element; *asnT*, asparagine-specific tRNA; SPI, Salmonella pathogenicity island; *inv*, invasion genes; *spa*, secretion protein antigen genes; *hil*, hyperinvasion locus; *valV*, valine-specific tRNA; CAG, cytotoxin-associated antigen; *glr*, glutamate racemase; VAP, virulence-associated protein; *vrl*, virulence-related locus; *serV*, serine-specific tRNA; *ssrA*, small stable RNA; TCP, toxin-coregulated pili; ACF, accessory colonization factor; *toxT*, transcriptional activator gene; *int*, integrase gene; *att*, attachment site.

which is involved in the regulation of virulence genes, may argue for a co-evolution of Pais and *leuX*.

The Pais of the uropathogenic *E. coli* strain J96 exhibit characteristics similar to those of strain 536 (see Table 1). Recent data indicate that the Pais of strain J96 are also located at tRNA genes, Pai I at *pheV* (64 units) and Pai II at *pheR* (94 units), both of which code for phenylalanine tRNAs (Donnenberg and Welch, 1995; Swenson *et al*., 1996). While the 3′ ends of the tRNA loci *selC* and *leuX* associated with Pais of strain 536 are recognized as attachment sites for the integration of the phages ϕR73 and P4, respectively (Inouye *et al*., 1991; Sun *et al*., 1991), the 3′ ends of *pheV* and *pheR* are identified as target sites for the integration of a conjugative transposon into the chromosomes of *Salmonella senftenberg* and *E. coli* (Hochhut *et al*., 1996; J. W. Lengeler, personal communication). Thus, the 3′ ends of tRNAs may generally act as sites for the integration of foreign DNA into bacterial genomes. Interestingly, cryptic integrase genes of phages ϕR73 and P4 are located in the vicinity of the respective *selC* and *leuX* tRNA genes on Pais I and II of strain 536. We therefore speculate that Pais (or parts of them) originate from integrated prophages, which may also be considered as 'Pai precursors' (Lee, 1996) or 'Pre-Pais'. In addition, sequences which show homology to origins of replication of plasmids, parts of IS elements and 'recombination hot spots' (Rhs), are located on the Pais of strains J96 and 536 (Blum *et al*., 1994; Donnenberg and Welch, 1995; Swenson *et al*., 1996; G. Blum, unpublished).

Another pathogenicity island of 35 kb has recently been identified on the chromosome of enteropathogenic *E. coli* (EPEC), and was found to induce an 'attaching and effacing' (AE) lesion following its transfer into *E. coli* K-12 (McDaniel *et al*., 1995; 1996). This region was therefore termed the 'locus of enterocyte effacing' (LEE). It carries genes encoding the adherence factor intimin (EaeA) as well as proteins involved in host signalling and type III secretion (Jarvis *et al*., 1995; see Table 1). Despite the fact that uropathogenic and enteropathogenic *E. coli* cause completely different infectious diseases, Pai I of the uropathogenic *E. coli* strain 536 and the LEE locus of the EPEC isolate E2348/69 are inserted at exactly the same positions into the tRNA locus *selC*. The LEE region also meets the requirements of the definition described in an earlier section. However, in contrast to the Pais of the uropathogenic *E. coli* strain, the insertion of the LEE-Pai into the chromosomes of EPEC and several enterohaemorrhagic *E. coli* (EHEC) strains seems to be stable.

Pathogenicity islands of other Gram-negative bacteria

Unstable Pais are located on the chromosomes of various species of the genus *Yersinia*, which comprises *Y. pestis*, the causative agent of plague, as well as the intestinal pathogens *Y. pseudotuberculosis* and *Y. enterocolitica*. The so-called 'high pathogenicity islands' (HPIs) of *Yersinia* can be lost at frequencies of 10^{-5} (Fetherston *et al*., 1992). The HPI region carries a pigmentation locus (*pgm*), coding for haemin storage (HMS) proteins, and genes encoding iron-uptake factors (see Table 1). The whole HPI region is present in *Y. pestis*, but only part of this region can be found in other *Yersinia* species. In *Y. pestis,* a 102 kb HPI locus is flanked by two IS*100* elements. Some strains have lost the whole 102 kb region because of a recombinational event occurring between the IS*100* elements. The fact that an open reading frame which shows homology to the *E. coli phoE* gene has been disrupted by one of the IS*100* elements (Fetherston and Perry, 1994) indicates that the deletable region, representing ≈2.3% of the *Y. pestis* genome, does not represent an autonomous genetic element, i.e. a pathogenicity island *per se*. The presence of IS elements, however, leads to a great flexibility of this region. Pathogenic and non-pathogenic *Y. enterocolitica* strains may differ by the presence of a 45 kb region within the 102 kb HPI, which harbours genes (*fyuA*, *irp2*) encoding iron-uptake factors, but not the *Y. pestis*-specific *hms* locus (Heesemann *et al*., 1993; Rakin *et al*., 1994). The 45 kb Pai of *Y. enterocolitica* strain Ye8081 is flanked on one side by the asparagine-specific tRNA gene *asnT*, and includes IS elements as well as other 'repeat sequences' (RSs) (Carniel *et al*., 1996). These elements seem to play a role in sequential deletions which may result in *Y. enterocolitica* variants that carry parts of the 45 kb HPIs. Interestingly, parts of the *Yersinia* islands, including the *fyuA* gene, are present in enteroaggregative *E. coli* (EAggEC) strains, supporting the argument for a spread of these sequences among different enterobacteria (J. Heesemann, personal communication). The capacity of *Yersinia* to lose and regain these Pais may represent a strategy by these bacteria to adapt themselves to different environments, such as soil, arthropods, and warm-blooded animals.

Chromosomal loci, termed 'Salmonella pathogenicity islands' (SPIs), have recently been described for *Salmonella typhimurium* (Mills *et al*., 1995; Shea *et al*., 1996). The gene products of SPI-I (see Table 1) are necessary for the invasion of *Salmonella* into epithelial cells. In contrast to SPI-I, SPI-II-specific gene products are essential if *S. typhimurium* is to survive within macrophages. While SPI-I does not show significant boundary sequences, SPI-II is located next to the valine-specific tRNA gene *valV* (M. Hensel and D. W. Holden, personal communication). The region next to *valV* in *E. coli* represents a recombinational hot spot in *E. coli* with two tRNA genes, one of which is missing in the *Salmonella* genome. According to the definition given in an earlier section, the SPIs are exceptional

Pais within Gram-negative bacteria because they are species- but not strain specific. Thus, SPI-I and SPI-II have been found in all *S. typhimurium* strains tested to date. Moreover, SPI-I but not SPI-II has also been detected in the species *Salmonella bongori*. Nevertheless, the SPI elements can be considered as Pais as they were not detected in the genome of related enterobacterial species, such as *Y. pestis, S. flexneri* and *E. coli*. It has therefore been speculated that SPI-specific DNA fragments were transmitted to *Salmonella* in the distant past (more than 100 million years ago), when *Salmonella* and other enterobacteria separated from a common ancestor (Lee, 1996).

An interaction between Pais and cholera toxin (CTX)-converting bacteriophages has recently been described for *V. cholerae* O1 and O139 strains. These strains carry a pathogenicity island of 45 kb, which comprises the *tcp–acf* gene cluster, the *toxT* gene involved in the regulation of the cholera toxin (*ctx*), and also an integrase gene (Kovach *et al*., 1996). Interestingly, the CTX phage uses the Pai-encoded Tcp adhesin as its receptor, providing evidence for co-evolution of the *V. cholerae*-specific Pai and the phage-associated *ctxAB* genes (Waldor and Mekalanos, 1996).

Another pathogenicity island, which meets the Pai criteria described earlier, has recently been identified in the genome of *Helicobacter pylori*, the causative agent of chronic gastritis and peptic-ulcer disease. *H. pylori* strains fall into two groups. The type I strains are associated with severe disease symptoms, while the disease caused by the type II strains is milder. The members of those two groups differ by the presence of the so-called *cag* (cytotoxin-associated antigen') pathogenicity island, which is 40 kb in size (Censini *et al*., 1996). Some Pai gene products show similarities to proteins of secretion pathways, including those of type III secretion. Indeed, this Pai was found to be necessary for the production of the vacuolating cytotoxin A (VagA), a major virulence factor of *H. pylori*. The *H. pylori* Pai is linked not to a tRNA gene but to the glutamate racemase gene (*glr*) and is flanked by 31 bp direct repeats. In some strains, the Pai is interrupted by the new insertion element, IS*605*, which is similar to an IS element found in Pais of *Dichelobacter nodosus* (see below). The IS*605* element and presumably the flanking sequences are involved in generating deletions of parts of the *cag* pathogenicity island, resulting in strains with intermediate virulence. It is suggested that the aquisition of this Pai is an important event in the evolution of *H. pylori* strains with higher virulence potencies.

Certain strains of *D. nodosus,* i.e. Gram-negative obligate anaerobic bacteria which cause ovine footroot (a necrotic disease in sheep) differ from each other with respect to their virulence potencies (Cheetham *et al*., 1995). The degrees of virulence of particular strains correspond with the presence of multiple copies of the so-called *vap* region, a DNA segment of 11.9 kb, which codes for virulence-associated proteins (Vaps) and the so-called virulence-related locus (*vrl*). The *vap* region of *D. nodosus* strain A198 is also located next to a tRNA locus, which is the serine-specific tRNA gene *serV*. One junction site of the *vrl* region is part of the 3′ end of a gene which is homologous to the *ssrA* gene of *E. coli*, encoding a small regulatory RNA. Moreover, *vap* gene products show similarities to plasmid-encoded proteins of other Gram-negative bacteria (Cheetham and Katz, 1995). Interestingly, a plasmid of *D. nodosus* which represents a circular form of the *vap* gene cluster has recently been described. It is capable of integrating into the chromosome, resulting in Pai formation (Billington *et al*., 1996).

Pathogenicity islands of Gram-positive bacteria

Pathogenicity islands have also been found in the genomes of Gram-positive pathogens. However, those Pais do not fulfil all the criteria of the Pai definition given earlier. They do not exhibit specific junction sites, such as direct repeats or tRNA loci, at their ends. Moreover, they do not carry 'mobility' genes and seem to be stably integrated into the genomes of the respective strains. Nevertheless, they encode important virulence factors and are specific for virulent strains of one species (e.g. *Clostridium difficile*) or for certain species of one genus (e.g. *Listeria monocytogenes, Streptococcus pyogenes*). The 19 kb Pai located in the genome of virulent strains of *C. difficile*, which is responsible for antibiotic-associated diarrhoea or its fatal form, pseudomembranous colitis, encodes the two high-molecular-weight toxins called TcdA (an enterotoxin) and TcdB (a cytotoxin) (Braun *et al*., 1997). Interestingly, a 115 bp fragment, which forms a 20 bp hairpin loop, is present only in non-pathogenic variants (not in pathogenic strains). It may represent the target site for the chromosomal integration of this pathogenicity island (Braun *et al*., 1997).

The 9.5 kb Pai of *L. monocytogenes,* which causes foodborne listeriosis, harbours five virulence genes. Except for *plcB*, all of these genes are regulated by the activator protein PrfA, the corresponding gene locus of which is physically linked to the virulence gene cluster (Lampidis *et al*., 1994). While all strains of the pathogenic species *L. monocytogenes* (pathogenic for humans) and *Listeria ivanovii* (pathogenic for animals) possess this pathogenicity island, strains of the non-pathogenic *Listeria* species *L. innocua*, *L. welshimeri* and *L. grayii* do not carry this specific DNA segment (Gouin *et al*., 1994; Kreft *et al*., 1995). Strains of the non-pathogenic species *L. seeligeri* harbour this Pai, although it is not properly expressed (J. Kreft, personal communication). In the *S. pyogenes* chromosome, but not in the genome of other species of the genus *Streptococcus*, the so-called *vir*-regulon, a Pai of up to 6 kb,

codes for a major regulator (*mga*), M and M-related proteins (*emm*, *mrp*, *enn*), and a C5b peptidase (*scpA*) (Podbielski *et al*., 1996).

Acquisition of Pais and microbial evolution

Point mutations, genomic rearrangements and horizontal gene transfer are driving forces in microbial evolution (Arber, 1993). While point mutations lead to 'slow' evolutionary development, the acquisition and excision of large genomic fragments quickly generate new variants of strains by means of 'genetic quantum leaps' (Falkow, 1996). Phages, plasmids and pathogenicity islands are involved in those processes of fast evolutionary movement. As Pais carry genes also found on phages and plasmids, the latter are also considered as 'Pai precursors' (Lee, 1996) or 'Pre-Pais'. The finding that Pais often exhibit features of both plasmids and phages may reflect the formation of co-integrates between both types of extrachromosomal elements, a process which has frequently been found in *Streptomyces* (Leblond and Decaris, 1994).

The acquisition of new genes following horizontal gene transfer either directly by transformation with naked DNA, transduction with phages, or the uptake of plasmids or chromosomal fragments by conjugation, results in the generation of new variants of pathogens. Drug-resistant strains of *Mycobacterium tuberculosis* and *Streptococcus pneumoniae*, the new *V. cholerae* serotype O139, enterohaemorrhagic *E. coli* of serotype O157, and the *Haemophilus influenzae* biotype aegypticus are examples of such newly discovered pathotypes (Achtman and Hakenbeck, 1992). Natural genetic competence, e.g. in the case of *H. influenzae* and *S. pneumoniae*, or in the case of species which do not exhibit natural competence, higher rates of mutation and recombination are prerequisites for the generation of new pathotypes. It has recently been shown that pathogenic enterobacterial strains, including those of *S. typhimurium* and *E. coli*, often carry defects in the *mutS* gene, which directs DNA-repair processes (LeClerc *et al*., 1996). The generation and acquisition of Pais may be directly associated with higher mutation rates and recombination efficiencies in pathogens (relative to those of commensal strains).

Following the transfer of phages, plasmids or Pais into new host cells, two genetic processes are of importance, i.e. stabilization of the new elements and optimal expression of the newly acquired genes. The high number of mutations, often leading to stop codons in the 'mobility' genes of Pais, such as integrases, origins of replication or IS elements, may reflect the process of genetic stabilization of the newly generated organisms to avoid the loss of the new genes, resulting in the conservation of an advantageous 'pathogenetic status quo' in certain strains. If Pai-specific genes are not incorporated into the regulatory network of the new host organism, they will not be properly expressed. This has been shown for the *Listeria* Pai, which is present in the genome of *Listeria seeligeri* but is not properly expressed because of an unusual promoter region of the *prf* regulator gene (J. Kreft, personal communication). Interestingly, Pais carry not only structural genes but also regulatory elements, as has been shown for adhesin regulators in uropathogenic *E. coli* strains (Morschhäuser *et al*., 1994) and for the *toxT* gene of the *V. cholerae* Pai (Kovach *et al*., 1996).

tRNA loci and other boundary regions of Pais

The boundary sequences of pathogenicity islands play an important role in their integration and/or excision. Thus, Pais are flanked by short direct repeats (e.g. uropathogenic *E. coli*, *D. nodosus, H. pylori*) which may act as targets for specific recombinases, thereby facilitating rearrangements. In addition, IS elements play a role in excision and presumably integration of Pais, as shown for *Y. pestis* (Fetherson and Perry, 1994), the Pai of which is flanked by two IS*100* elements, and for the *E. coli* aerobactin gene, which is flanked by IS*1* elements that allow the determinant to switch between the chromosome and plasmids (Neilands, 1992).

Interestingly, pathogenicity islands of many species (e.g. *E. coli, S. typhimurium, D. nodosus*) are located in the vicinity of tRNA genes. The finding that the pathogenicity islands of uropathogenic and enteropathogenic *E. coli,* coding for completely different virulence factors, are located in *selC* tRNA genes at identical base pairs, strongly argues for the specificity of recombination processes leading to integration of Pais (Blum *et al*., 1994; Mc Daniel *et al*., 1995). tRNA loci were found to be the targets for the integration of plasmids and phages into chromosomes of various bacterial species, such as *E. coli* (Inouye *et al*., 1991), *Streptomyces* spp. (Alegre *et al*., 1994) and *Pseudomonas* (Hayashi *et al*., 1993), and even of eukaryotes (Marschalek *et al*., 1989). Moreover, phages present in different microorganisms carry tRNA-specific sequences (Campbell, 1992). As tRNAs are also involved in the expression of certain genes, these loci may play an additional role in the regulation of Pai- and phage-specific genes (Ritter *et al*., 1995). Thus it can be concluded that tRNA genes act as genomic landmarks for foreign genetic determinants and regulatory elements. The recent findings with respect to the association of pathogenicity islands and tRNA genes give ample evidence for the reasoning that tRNA loci play a key role in the evolution of microbial pathogens.

Deletions as processes of adaptation and genome plasticity

From an evolutionary point of view, not only gene

acquisition but also the formation of deletions represents a major principle of genome plasticity (Ott, 1993). The excision phenomenon has frequently been analysed in *Streptomyces* spp., where large deletions comprising DNA of up to 800 kb occur (Birch *et al*., 1991; Leblond and Decaris, 1994). The deleted parts of the chromosomes are often flanked by either perfect or non-perfect repeats, which may play a role in illegitimate recombination. The permanent generation of new genetic variants due to deletions may be an advantage for *Streptomyces* in the adaptation to the environment. In other species, such as *Bacillus subtilis* or cyanobacteria, excisions of DNA fragments which interrupt coding regions have been described (Haselkorn, 1992). The excision of 42 kb of the *B. subtilis* genome is a very good example of 'developmental' deletions because the interrupted sporulation (*spo*) genes are fused following this deletion, leading to a functional, active sporulation determinant (Stragier *et al*., 1989). The integration and excision of plasmids into and from chromosomes have also been described for *Streptomyces* spp. (Brasch *et al*., 1993), *Acinetobacter calcoaceticus* (Ka and Tiedje, 1994), *S. flexneri* and enteroinvasive *E. coli* (Colonna *et al*., 1995). It is speculated that the site-specific excisions (e.g. uropathogenic *E. coli*) or sequential deletions (*Y. pestis*, *Y. enterocolitica, H. pylori*) of pathogenicity islands or other genetic segments from chromosomes of bacterial pathogens may also play a role in the bacterial adaptation to different environments (Rajahumar *et al*., 1996).

How to identify new Pais?

While the first Pais were described only a few years ago, the occurrence of an increasing number of such elements has been reported recently (see Table 1). More Pais are likely to be identified in many different pathogens in the near future. Molecular epidemiology will certainly lead to the description of new variants of pathogenic species, which will form the basis for the detection of new pathogenicity islands. The application of new techniques, such as signature tagging, subtraction hybridization, and differential display of mRNA following RT-PCR, all of which have been used successfully to identify new pathogenicity islands (Shea *et al*., 1996; Censini *et al*., 1996; Maheiras *et al*., 1996), will certainly make it possible to identify and characterize unknown Pais. In addition, new cloning techniques (Bloch *et al*., 1996) and DNA sequence data on Pai-specific regions will provide new information in the near future. These data will broaden our knowledge to allow further speculation on the evolutional mechanisms that form the basis for the creation of new pathogens. It is also speculated that genetic elements exist which are similar to Pais and play a role in secretion ('secretion islands'), resistance to antibiotics ('resistance islands') or in the physiology of microorganisms ('metabolic islands'). However, the genetic structures of these putative islands remain to be determined.

Acknowledgements

The authors thank many colleagues, too numerous to mention, who contributed to this review in the form of unpublished data or forthcoming manuscripts. Our own work on this topic is supported by the Deutsche Forschungsgemeinschaft (Ha 1434/8–2) and the Fond der Chemischen Industrie.

References

Achtman, M., and Hakenbeck, R. (1992) Recent developments regarding the evolution of pathogenic bacteria. In *Molecular Biology of Bacterial Infection: Current Status and Future Perspective*. Vol. 49. Hormaeche, C.E. (ed.). Cambridge: SGM/Cambridge University of Press, pp. 13–31.

Alegre, M.-T., Cournoyer, B., Mesas, J.-M., Guérineau, M., Normand, P., and Pernodet, J.-L. (1994) Cloning of *Frankia* species putative $tRNA^{Pro}$ genes and their efficiacy for pSAM2 site-specific integration in *Streptomyces lividans. Appl Environ Microbiol* **60:** 4279–4283.

Arber, W. (1993) Evolution of prokaryotic genomes. *Gene* **135:** 49–56.

Billington, S.J., Sinistaj, M., Cheetham, B.F., Moses, E.K., Katz, M.E., and Rood, J.I. (1996) Identification of a native *Dichelobacter nodosus* plasmid and implications for the evolution of the *vap* regions. *Gene* **172:** 111–118.

Birch, A., Häusler, A., Rüttener, C., and Hüttner, R. (1991) Chromosomal deletion and rearrangements in *Streptomyces glaucescens. J Bacteriol* **173:** 3531–3538.

Bloch, C.A., and Rode, C.K. (1996) Pathogenicity island evaluation in *Escherichia coli* K1 by crossing with laboratory strain K-12. *Infect Immun* **64:** 3218–3223.

Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschäpe, H., and Hacker, J. (1994) Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* **62:** 606–614.

Blum, G., Falbo, B., Caprioli, A., and Hacker, J. (1995) Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and α-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. *FEMS Microb Lett* **126:** 189–196.

Brasch, M.A., Pettis, G.S., Lee, S.C., and Cohen, S.N. (1993) Localization and nucleotide sequences of genes mediating site-specific recombination of the SLP1 element in *Streptomyces lividans. J Bacteriol* **175:** 3067–3074.

Braun, V., Hundsberger, T., Leukel, P. Sauerborn, M., and Eichel-Streiber, C. (1997) Definition of the single integration site of the pathogenicity locus in *Clostridium difficile. Gene*, in press.

Campbell, A.M. (1992) Chromosomal insertion sites for phages and plasmids. *J Bacteriol* **174:** 7495–7499.

Carniel, E., Guilvout, I., and Prentice, M. (1996) Characterization of a large chromosomal 'High Pathogenicity Island' in biotype 1B *Yersinia enterocolitica. J Bacteriol* **178:** 6743–6751.

Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., Rappuoli, R., and Covacci, A. (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA* **93:** 14648–14653.

Cheetham, B.F., and Katz, M.E. (1995) A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol Microbiol* **18:** 201–208.

Cheetham, B.F. Tattersall, D.B., Bloomfield, G.A., Rood, J.I., and Katz, M.E. (1995) Identification of a gene encoding a bacteriophage-related integrase in a *vap* region of the *Dichelobacter nodosus* genome. *Gene* **162:** 53–58.

Colonna, B., Casalino, M., Fradiani, P.A., Zagaglia, C., Naitza, S., Leoni, L., Prosseda, G., Coppo, A., Gherlardini, P., and Nicoletti, M. (1995) H-NS regulation of virulence gene expression in enteroinvasive *Escherichia coli* harboring the virulence plasmid integrated into the host chromosome. *J Bacteriol* **177:** 4703–4712.

Donnenberg, M.S., and Welch, R.A. (1995) Virulence determinants of *uropathogenic Escherichia coli*. In *Urinary Tract Infections: Molecular Pathogenesis and Clinical Management*. Mobley, H.L.T., and Warren, J.W. (eds). Washington DC: American Society for Microbiology Press, pp. 135–174.

Falkow, S. (1996) The evolution of pathogenicity in *Escherichia*, *Shigella* and *Salmonella*. In Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 2723–2729.

Fetherston, J.D., Schuetze, P., and Perry, R.D. (1992) Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. *Mol Microbiol* **6:** 2693–2704.

Fetherston, J.D., and Perry, R.D. (1994) The pigmentation locus of *Yersinia pestis* KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. *Mol Microbiol* **13:** 697–708.

Galan, J.E., and Sansonetti, F. (1996) Molecular and cellular bases of *Salmonella* and *Shigella* interactions with host cells. In Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 2757–2773.

Gouin, E., Mengaud, J., and Cossart, P. (1994) The virulence gene cluster of *Listeria monocytogenes* is also present in *Listeria ivanovii*, an animal pathogen, and *Listeria seeligeri*, a non-pathogenic species. *Infect Immun* **62:** 3550–3553.

Groisman, A.G., and Ochman, H. (1996) Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87:** 791–794.

Hacker, J., Knapp, S., and Goebel, W. (1983) Spontaneous deletions and flanking regions of the chromosomal inherited hemolysin determinant of an *Escherichia coli* O6 strain. *J Bacteriol* **154:** 1145–1152.

Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R., and Goebel, W. (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vivo* and *in vitro* in various extraintestinal *Escherichia coli* isolates. *Microb Pathogen* **8:** 213–225.

Haselkorn, R. (1992) Developmentally regulated gene rearrangements in prokaryotes. *Annu Rev Genet* **26:** 113–130.

Hayashi, T., Matsumoto, H., Ohnishi, M., and Terawaki, Y. (1993) Molecular analysis of a cytotoxin-converting phage, ϕCTX, of *Pseudomonas aeruginosa*: structure of the *att*P–*cos*–*ctx* region and integration into the serine tRNA gene. *Mol Microbiol* **7:** 657–667.

Heesemann, J., Hantke, K., Vocke, T., Saken, E., Rakin, A., Stojilijkovic, I., and Berner, R. (1993) Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65 000 Da and pesticin sensitivity. *Mol Microbiol* **3:** 397–408.

Hochhut, B., Schmid, K., Jahreis, K., and Lengeler, J.W. (1996) Analysis of scr-94, a conjugative transposon in enterobacteria. In *Abstracts of the Annual Meeting of the VAAM, 1996. Biospektrum,* Sonderausgabe PE130. Meyer, O. (ed.). Heidelberg: Spektrum, p. 121.

Inouye, S., Sunshine, M.G., Six, E.W., and Inouye, M. (1991) Retronphage ϕR73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. *Science* **252:** 969–971.

Jarvis, K.G., Girón, J.A., Jerse, A.E., McDaniel, T.K., Donnenberg, M.S., and Kaper, J.B. (1995) Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. *Proc Natl Acad Sci USA* **92:** 7996–8000.

Ka, J.O., and Tiedje, J.M. *(1994)* Integration and excision of a 2,4-dichlorophenoxyacetic acid-degradative plasmid in *Acaligenes paradoxus* and evidence of its natural intergenic transfer. *J Bacteriol* **176:** 5284–5289.

Knapp, S., Then, I., Wels, W., Michel, G., Tschäpe, H., Hacker, J., and Goebel, W. (1985) Analysis of the flanking regions from different hemolysin determinants of *Escherichia coli. Mol Gen Genet* **200:** 385–392.

Knapp, S., Hacker, J., Jarchau, T., and Goebel, W. (1986) Large unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* strain 536. *J Bacteriol* **168:** 22–30.

Kovach, M.E., Shefter, M.D., and Peterson, K.M. (1996) A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio cholerae. Microbiology* **142:** 2165–2174.

Kreft, J., Dumbsky, M., and Theiss, S. (1995) The actin-polymerization protein from *Listeria ivanovii* is a large repeat protein which shows only limited amino acid sequence homology to ActA from *Listeria monocytogenes. FEMS Microbiol Lett* **126:** 113–122.

Lampidis, R., Gross, R., Sokolovic, Z., Goebel, W., and Kreft, J. (1994) The virulence regulator protein of *Listeria ivanovii* is highly homologous to PrfA from *Listeria monocytogenes* and both belong to the Crp–Fnr family of transcription regulators. *Mol Microbiol* **13:** 141–151.

Leblond, P., and Decaris, B. (1994) New insights into the genetic instability of Streptomyces. *FEMS Microbiol Lett* **123:** 225–232.

Le Clerc, J.E., Li, B., Payne, W.L., and Cebula, T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274:** 1208–1211.

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

Low, D., David, V., Lark, D., Schoolnik, G., and Falkow, S.

(1984) Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infections. *Infect Immun* **43:** 353–358.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., and Kaper, J.B. (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci USA* **92:** 1664–1668.

McDaniel, T.K., Ying Kang, D., and Kaper, J.B. (1996) The EPEC locus of enterocyte effacement: five more *sep* genes and genetic reconstruction of the attaching and effacing phenotype in *E. coli* K-12. In *Abstracts of the 96th General Meeting of the American Society for Microbiology, 1996*. Washington, DC: American Society for Microbiology, p. 170.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.V., and Stover, C.K. (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178:** 1274–1284.

Marschalek, R., Brechner, T., Amon-Böhm, E., and Dingermann, T. (1989) Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. *Science* **244:** 1493–1496.

Mills, D.M., Bajaj, V., and Lee, C.A. (1995) A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol Microbiol* **15:** 749–759.

Morschhäuser, J., Vetter, V., Emödy, L., and Hacker, J. (1994) Adhesin regulatory genes as parts of large, unstable DNA regions of pathogenic *Escherichia coli*: cross-talk between different adhesin gene clusters. *Mol Microbiol* **11:** 555–566.

Mühldorfer, I., and Hacker, J. (1994) Genetic aspects of *Escherichia coli* virulence. *Microb Pathogen* **16:** 171–181.

Neilands, J.B. (1992) Mechanisms and regulation of synthesis of aerobactin in *Escherichia coli* K-12 (pColV-K30). *Can J Microbiol* **38:** 728–733.

Ott, M. (1993) Dynamics of the bacterial genome: deletions and integrations as mechanisms of bacterial virulence modulation. *Zbl Bakt* **278:** 457–468.

Podbielski, A., Woischnik, M., Pohl, B., and Schmidt, K.H. (1996) What is the size of the group A streptococcal *vir* regulon? The Mga regulator affects expression of secreted and surface virulence factors. *Med Microbiol Immunol*, in press.

Rajahumar, K., Sasakawa, C., and Adler, B. (1996) A spontaneous 99 kb chromosomal deletion results in multi-antibiotic susceptibility and an attenuation of contact hemolysis in *Shigella flexneri* 2a. *J Med Microbiol* **45:** 64–75.

Rakin, A., Saken, E., Harmsen, D., and Heesemann, J. (1994) The pesticin receptor of *Yersinia enterocolitica*: a novel virulence factor with dual function. *Mol Microbiol* **13:** 253–263.

Ritter, A., Blum, G., Emödy, L., Kerenyi, M., Böck, A., Neuhierl, B., Rabsch, W., Scheutz, F., and Hacker, J. (1995) tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. *Mol Microbiol* **17:** 109–121.

Salyers, A.A., and Whitt, D.D. (1994) *Bacterial Pathogenesis: A Molecular Approach*. Washington, DC: American Society for Microbiology Press.

Shea, J.E., Hensel, M., Gleeson, C., and Holden, D.W. (1996) Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. *Proc Natl Acad Sci USA* **93:** 2593–2597.

Stragier, P., Kunkel, B., Kroos, L., and Losick, R. (1989) Chromosmal rearrangements generating a composite gene for the developmental transcription factor. *Science* **243:** 507–512.

Sun, J., Inouye, M., and Inouye, S. (1991) Association of a retroelement with a P4-like cryptic prophage (retronphage ϕR73) integrated into the selenocystyl tRNA gene of *Escherichia coli*. *J Bacteriol* **173:** 4171–4181.

Susa, M., Kreft, B., Wasenauer, G., Ritter, A., Hacker, J., and Marre, R. (1996) Influence of cloned tRNA genes from a uropathogenic *Escherichia coli* strain on adherence to primary human renal tubular epithelial cells and nephropathogenicity in rats. *Infect Immun* **64:** 5390–5394.

Swenson, D.L., Bukanov, N.O., Berg, D.E., and Welch, R.A. (1996) Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. *Infect Immun* **64:** 3736–3743.

Waldor, M.K., and Mekalanos, J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272:** 1910–1914.