MicroReview

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

J. Hacker,1* G. Blum-Oehler,1 I. Mühldorfer1 and H. Tschäpe2
1Institut für Molekulare Infektionsbiologie, Röntgenring 11, 97070 Würzburg, Germany.
2Robert-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.

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 uropathogenesis 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; Mühl dorfer 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 Pains with sizes of 70–190 kb (see Table 1). The Pains 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 Pains 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 Pains 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 Pai 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 Pains are part of phage-attachment sites located at the 3’ end of tRNA genes. Thus, Pai I is flanked by the seIC gene (82 units), encoding the selenocysteine-specific tRNA, while Pai II is associated with leuX, which codes for a minor leucyl tRNA 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,
**Table 1. Pathogenicity islands (Pais) of Gram-negative bacteria.**

<table>
<thead>
<tr>
<th>Pathogenic microorganism</th>
<th>Pai designation</th>
<th>Relevant genes</th>
<th>G+C content (%)</th>
<th>Size (kb)</th>
<th>Location (units)</th>
<th>Boundary sequence</th>
<th>Associated tRNA</th>
<th>Deletion frequency</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli (536)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uropathogenic</td>
<td>Pai I</td>
<td>hly</td>
<td>51/41</td>
<td>70</td>
<td>82</td>
<td>DR 16bp</td>
<td>selC</td>
<td>10⁻⁴–10⁻⁵</td>
<td>Blum et al. (1994)</td>
</tr>
<tr>
<td>Uropathogenic</td>
<td>Pai II</td>
<td>hly prel</td>
<td>51/41</td>
<td>190</td>
<td>97</td>
<td>DR 18bp</td>
<td>leuX</td>
<td>10⁻⁴–10⁻⁵</td>
<td>Blum et al. (1994)</td>
</tr>
<tr>
<td><strong>E. coli (J96)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uropathogenic</td>
<td>Pai IV</td>
<td>hly papa</td>
<td>51/41</td>
<td>170</td>
<td>64</td>
<td>–</td>
<td>prf</td>
<td>?</td>
<td>Swenson et al. (1996)</td>
</tr>
<tr>
<td>Uropathogenic</td>
<td>Pai V</td>
<td>hly prel cnfl</td>
<td>51/41</td>
<td>110</td>
<td>94</td>
<td>DR 135bp</td>
<td>pheR</td>
<td>10⁻⁵</td>
<td>Blum et al. (1995); Swenson et al. (1996)</td>
</tr>
<tr>
<td><strong>E. coli (E2348/69)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Y. pestis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Y. enterocolitica</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPI</td>
<td></td>
<td>fyUA ip2 (ip1)</td>
<td>46–50/56</td>
<td>46</td>
<td>45</td>
<td>–</td>
<td>asnT</td>
<td>10⁻⁶</td>
<td>Carniel et al. (1996)</td>
</tr>
<tr>
<td><strong>S. typhimurium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPI I</td>
<td></td>
<td>inv spa hil</td>
<td>52/40–47</td>
<td>40</td>
<td>60.3</td>
<td>–</td>
<td>–</td>
<td>Stable/unstable⁴</td>
<td>Mills et al. (1995)</td>
</tr>
<tr>
<td><strong>S. typhimurium</strong></td>
<td></td>
<td>Type III secretion genes, two-component genes</td>
<td>52/40–47</td>
<td>40</td>
<td>30.7</td>
<td>–</td>
<td>valV</td>
<td>Stable</td>
<td>Shea et al. (1996); M. Hensel &amp; D. W. Holden, personal communication</td>
</tr>
<tr>
<td><strong>H. pylori</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cag Pai</td>
<td></td>
<td>cagA–T</td>
<td>38–45/35</td>
<td>40</td>
<td>40</td>
<td>DR 31bp</td>
<td>gfr</td>
<td>Unstable⁴</td>
<td>Censini et al. (1997)</td>
</tr>
<tr>
<td><strong>D. nodosus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D. nodosus</strong></td>
<td></td>
<td>Vrl region</td>
<td>49/57</td>
<td>27</td>
<td>–</td>
<td>–</td>
<td>ssrA</td>
<td>?</td>
<td>Billington et al. (1996)</td>
</tr>
<tr>
<td><strong>V. cholerae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCP–ACF element</td>
<td></td>
<td>acf tcp toxT int</td>
<td>–</td>
<td>45</td>
<td>–</td>
<td>att sites</td>
<td>20bp</td>
<td>Stable</td>
<td>Kovach et al. (1996)</td>
</tr>
</tbody>
</table>

**Abbreviations:** hly, haemolysin gene cluster; prel, 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; cnfl, cytotoxop-hemolysin-coding 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, haem storage protein gene HFRS; fyUA, fimbria expression factor A; ip, 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; gfr, glutamate racemase; VAP, virulence-associated protein; vRL, virulence-related locus; servV, serine-specific tRNA; ssrA, small stable RNA; TCP, toxin-coregulated pilus; ACF, accessory colonization factor; toxT, transcriptional activator gene; int, integrase gene; att, attachment site.

**Notes:**

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).

c. VAP, virulence-associated protein; vRL, virulence-related locus; servV, serine-specific tRNA; ssrA, small stable RNA; TCP, toxin-coregulated pilus; ACF, accessory colonization factor; toxT, transcriptional activator gene; int, integrase gene; att, attachment site.

**References:**

- Blum et al. (1994)
- Blum et al. (1995)
- Swenson et al. (1996)
- McDaniel et al. (1995)
- Fetherston and Perry (1994)
- Carniel et al. (1996)
- Mills et al. (1995)
- Censini et al. (1997)
- Cechtham et al. (1995)
- Billington et al. (1996)
- Kovach et al. (1996)
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 R73 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 Pai I and II of strain 536. We therefore speculate that Pai (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 Pai 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 IS100 elements. Some strains have lost the whole 102 kb region because of a recombinational event occurring between the IS100 elements. The fact that an open reading frame which shows homology to the *E. coli* *phoE* gene has been disrupted by one of the IS100 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*, *ipr2*) 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, theCTX 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 (cytotxin-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, IS605, which is similar to an IS element found in Pais of Dichelobacter nodosus (see below). The IS605 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 acquisition 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 fulfill 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. grayi 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 (emn, mnp, 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, enterohemorrhagic 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 (Fetherston and Perry, 1994), the Pai of which is flanked by two IS100 elements, and for the E. coli aerobactin gene, which is flanked by IS1 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; McDaniel 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 developmentaI 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 evolutionary 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**


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


