

## Microreview

# Pathogenicity islands: a molecular toolbox for bacterial virulence

Ohad Gal-Mor and B. Brett Finlay\*

Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada.

### Summary

**Pathogenicity islands (PAIs) are distinct genetic elements on the chromosomes of a large number of bacterial pathogens. PAIs encode various virulence factors and are normally absent from non-pathogenic strains of the same or closely related species. PAIs are considered to be a subclass of genomic islands that are acquired by horizontal gene transfer via transduction, conjugation and transformation, and provide ‘quantum leaps’ in microbial evolution. Data based on numerous sequenced bacterial genomes demonstrate that PAIs are present in a wide range of both Gram-positive and Gram-negative bacterial pathogens of humans, animals and plants. Recent research focused on PAIs has not only led to the identification of many novel virulence factors used by these species during infection of their respective hosts, but also dramatically changed our way of thinking about the evolution of bacterial virulence.**

Nothing in biology makes sense except in the light of evolution.

Theodosius Dobzhansky (1900–1975)

Pathogens are characterized by defined differences from their non-pathogenic relatives: they have evolved the ability to cause disease in other organisms. In the following discussion, it is important to acknowledge that every microorganism has adapted to a particular niche, and pathogens are no exception. Pathogenicity represents another bacterial lifestyle, with the host serving merely as an additional ecological niche. As such, the evolutionary forces driving the adaptation of microorganisms to environmental niches function the same way in the evolution of pathogens. Thus, as it is true for all microbial genomes,

bacterial pathogens have evolved by three major processes: (i) modification of existing genes, (ii) loss of genes no longer under selection, and (iii) gain of genes that confer benefit in their current ecological niche. In this review we will discuss various aspects of the third process, by focusing on the molecular and evolutionary mechanisms that lead to the acquisition and formation of pathogenicity islands and their contribution to bacterial virulence.

### Pathogenicity islands

The first published complete microbial genome sequence of *Haemophilus influenzae* in 1995 (Fleischmann *et al.*, 1995) initiated a new era in bacterial pathogen evolutionary research. To date, the complete sequence of more than 300 bacterial genomes has been published and the sequencing of more than 940 other bacterial genomes is currently ongoing (<http://www.genomesonline.org/>). Comprehensive analyses of the sequence data have suggested that the genes encoding virulence functions, in many pathogens, were somehow different from other genes in the chromosome. Differences often include changes in overall nucleotide composition, codon usage bias, association with mobile genetic elements, and linkage with tRNA genes. Analyses have also shown that bacterial genomes consist of a mosaic DNA architecture of various guanine and cytosine (G + C) content segments. The major portion of the genomic sequence (normally 70%–80%) is of homogeneous G + C content, which is typical to each bacterial species. This conserved and stable ‘core genome’ contains the genetic information that is required for essential cellular functions. In contrast, the rest of the genome (20%–30%) carries incongruous regions of large foreign DNA segments with distinct G + C content, which are scattered throughout the genome forming an ‘flexible gene pool’ (Hacker and Kaper, 2000).

This flexible gene pool includes mobile (or formerly mobile) genetic elements [such as insertion sequences (ISs), transposons, integrons, plasmid and prophages] as well as large unstable regions that have been designated ‘genomic islands’ (GEIs). These GEIs typically encode different genes associated with various symbiotic or

Received 25 May, 2006; revised 10 July, 2006; accepted 11 July, 2006. \*For correspondence. E-mail bfinlay@interchange.ubc.ca; Tel. (+1) 604 822 2210; Fax (+1) 604 822 9830.

pathogenic functions of bacteria. GEIs that encode virulence factors of pathogenic bacteria are referred to as 'pathogenicity islands' (PAIs). Similarly, depending on the function encoded by the GEI, they may also be called symbiosis, metabolic, or resistance islands (Hacker and Kaper, 2000).

The term PAI was coined by Hacker *et al.* to describe two large unstable regions on the chromosome of uropathogenic *Escherichia coli* (UPEC) (Hacker *et al.*, 1990). Currently, this term is commonly used to describe regions in the genomes of certain pathogens that are absent in the non-pathogenic strains of the same or closely related species and that contain large continuous blocks of virulence genes. PAIs recently characterized in a wide range of bacterial pathogens have not only led to the identification of many virulence factors used by these species, but also changed our way of thinking about the evolution of bacterial pathogenicity.

### Properties of PAIs

Although PAIs are diverse in structure and function, many of them share several common features. PAIs carry one or more virulence-associated gene and can occupy relatively large regions of the chromosome, ranging from 10 kb to more than 100 kb. Some strains also harbour smaller pieces of DNA (1–10 kb) that have been termed 'pathogenicity islets'. Often, PAIs have a G + C content and a codon usage that differs from that of the rest of the core genome. PAIs are frequently flanked on one side by a tRNA gene and often by direct repeat (DR) sequences. Additionally, PAIs commonly possess genes encoding factors that are involved in genetic mobility, such as integrases, transposases, phage genes and origins of replication. Some of the PAIs characteristics and specific examples are discussed below to illustrate some of these properties:

#### Virulence genes

PAIs encode many different functions, which largely depend on the environmental context in which the bacterium lives. The genetic repertoire found within PAIs can be functionally divided into several groups. The most common groups are: (i) Adherence factors, such as P-related pili, S-fimbriae, the *Vibrio cholerae* toxin-coregulated pilus (TCP) and intimin. These components enable bacteria to attach to host surfaces and to facilitate the infection process. (ii) Siderophores like the yersinia-bactin or the aerobactin used to deliver the essential element iron, barely soluble under aerobic or neutral pH conditions, into microbial cells. (iii) Exotoxins such as  $\alpha$ -haemolysin, enterotoxins and the pertussis toxin, which destroy or affect the function of eukaryotic cells. (iv) Inva-

sion genes which mediate bacterial entry into eukaryotic cells such as the *inv* genes of *Salmonella* spp. (v) Type III and IV secretion systems, conserved organelles that deliver bacterial effector proteins capable of modulating host functions into host cells. Type III secretion systems (T3SS) can be found on PAIs of many pathogens including *Salmonella* [pathogenicity island (SPI) 1] and SPI 2, *Shigella*, *Yersinia*, *Citrobacter*, and in different species of the *E. coli* phylogenetic lineages. Examples of PAIs harbouring type IV secretion systems (T4SS) include *Agrobacterium tumefaciens*, *Legionella pneumophila*, *Helicobacter pylori*, *Bordetella pertussis*, *Bartonella* spp. and *Brucella* spp.

#### tRNA genes

Some tRNA genes represent hot spots for the integration of foreign DNA, including PAIs. The 3' end of tRNA genes is frequently identical to the attachment sites of bacteriophages, and therefore function as preferred target sites for integration of certain plasmids and phages in various bacteria (Reiter *et al.*, 1989). The first examples showing that PAIs are inserted into tRNA-specific loci were the PAI I and PAI II of the UPEC 536. PAI II is integrated into the locus of the leucine tRNA gene (*leuX*), while PAI I was shown to be inserted into the tRNA gene of selenocysteine (*selC*) (Blum *et al.*, 1994).

Indeed, many of the PAIs in enterobacteria appear linked to either *selC* or one of the two genes of phenylalanine-tRNA, *pheV* or *pheU*. This rule, however, does not hold true in other pathogens, in which insertion of PAIs can be found in a similar frequency at other tRNA sites.

In some cases, the integration of PAI can occur into different chromosomal loci (copies) of a certain tRNA gene. The so-called high pathogenicity island (HPI) of the genus *Yersinia* can be integrated into any of the three chromosomal copies of the asparagine tRNA gene (*asn*) and is able to 'jump' from one locus to another on the same chromosome (Carniel *et al.*, 1996).

#### Functional mobility genes

PAIs often carry cryptic or functional mobility genes such as phage-like integrase genes, termed *int*, or genes for transposases. The virulence-associated region (*vap*) of *Dichelobacter nodosus*, for example, contains an open reading frame (ORF) with high level of sequence similarity to the integrase of an *E. coli* phage (Cheetham and Katz, 1995); and the HPI of *Yersinia* carries a phage P4 integrase homologue (Buchrieser *et al.*, 1998). Other PAIs contain genes that are similar to phages integrase and transposons resolvase genes. These gene products are involved in insertion and excision of the DNA regions by

recombination between flanking DRs, IS elements, or between regions of homologous sequences. Therefore, a subset of PAIs of some pathogens like *H. pylori*, *Yersinia* spp. and UPEC are genetically unstable and may be spontaneously deleted at frequencies of  $10^{-4}$ – $10^{-6}$  (Fetherston *et al.*, 1992; Blum *et al.*, 1994; Hacker *et al.*, 1997). In contrast, the PAIs of *Salmonella* and intestinal *E. coli* appear to be permanently integrated into the chromosome.

#### Direct repeat

Some PAIs are flanked by DR sequences, defined as DNA sequences usually between 16 and 20 bp with perfect or nearly perfect sequence repetition. The repeats are frequently homologous to phage attachment sites and are probably generated during the integration of mobile genetic elements into the host chromosome, via site-specific recombination resulting in the duplication of the integration site. DRs act as recognition sequences for enzymes involved in excision of mobile genetic elements and therefore can contribute to genomic instability of the island (Hacker *et al.*, 1997).

#### Insertion sequences

Insertion sequences are small mobile genetic elements, capable of transposing within and between prokaryotic genomes. IS provide sites of inverted repeats, at which homologous recombination may occur, and therefore can mediate the incorporation of mobile genetic elements into the chromosome, resulting in PAIs but can also contribute to PAIs excision or instability (Hacker *et al.*, 1997). For example, in *Yersinia pestis*, the *pgm* PAI is flanked by DRs of IS100 (Fetherston *et al.*, 1992). This element has about 30 copies in the *Yersinia* genome and can mediate integration of plasmids into the chromosome. A different PAI in the *Yersinia* chromosome, HPI, possesses several repeated sequences such as IS1328, IS1400 and RS3 (Carniel *et al.*, 1996). In *H. pylori* rearrangement of the *cag* island is known to be mediated by IS605 (Odenbreit and Haas, 2002).

#### Acquisition of PAIs

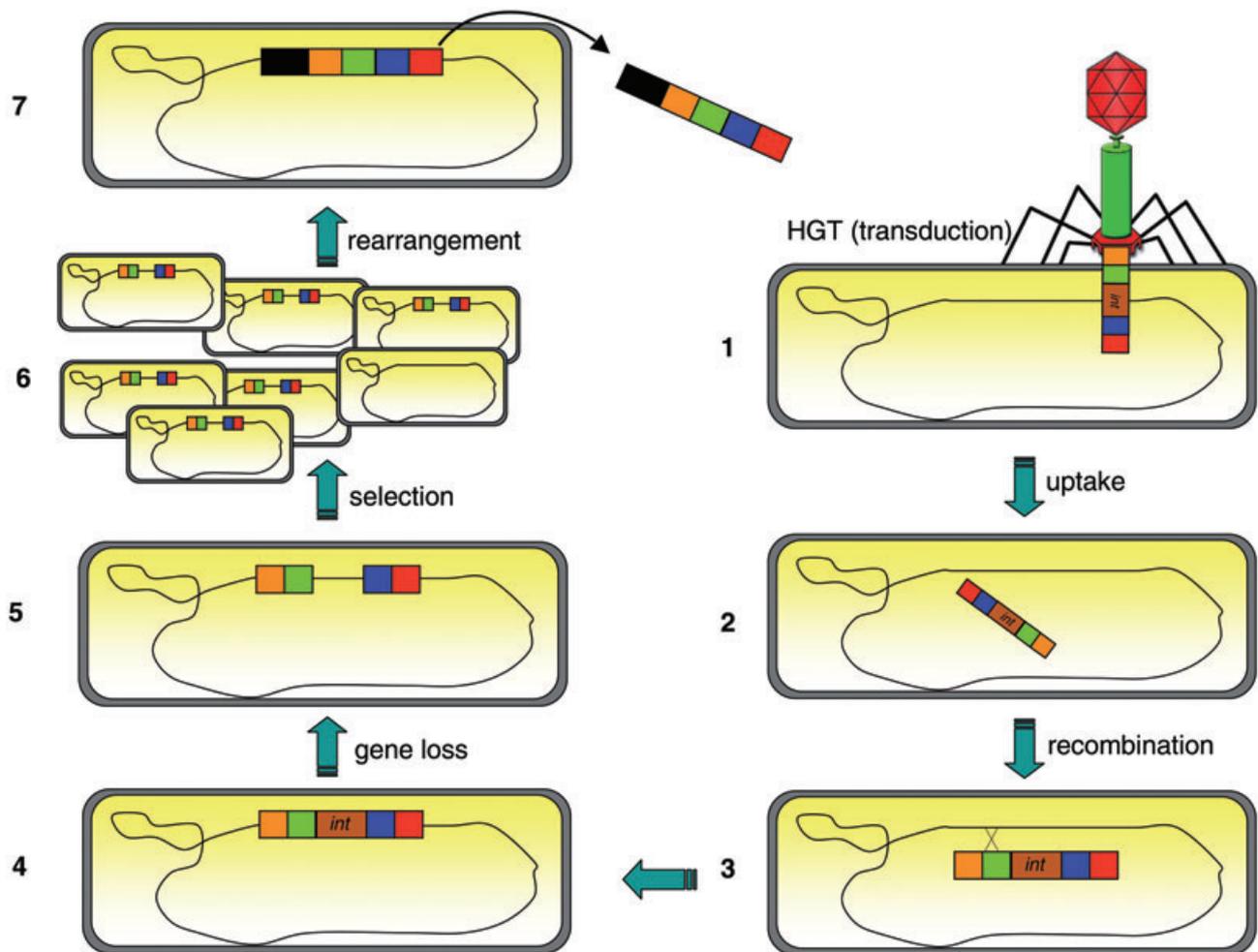
Horizontal gene transfer (HGT) is defined as the transfer of genetic material between bacterial cells uncoupled with cell division (Lawrence, 2005). Before the era of genomics, it was known that bacteria acquired foreign DNA such as plasmids and phages. However, exchange of genetic information by HGT and homologous recombination has been greatly underestimated in both quantity and quality (Doolittle, 1999). The conservative view that the evolution of prokaryotes occurs by clonal divergence and periodic

selection is now being broadened and instances of HGT can no longer be dismissed as 'exceptions that prove the rule'. Instead, our current view perceives genetic exchange by HGT as a major and continuing force in microbial evolution. This notion was strikingly demonstrated in the comparison of pathogenic and non-pathogenic strains of *E. coli* that share only 40% of their genome (Welch *et al.*, 2002), indicating that the rest has been acquired horizontally during the evolution.

There are three principal mechanisms that facilitate HGT in prokaryotes: natural transformation; conjugation; and transduction (reviewed in Jain *et al.*, 2002). Transformation is the process whereby prokaryotes take up free DNA from their surroundings. Conjugation, involves transfer of conjugative plasmids through a designated tube-like structure (a pilus, which represents a subset of the T4SS family) from a donor bacterium to a recipient cell that can be distantly related prokaryotic species. Transduction is the movement of genes from one prokaryotic species to another via viruses.

According to the model of Hacker and Carniel (Hacker and Carniel, 2001), one can envision five steps that lead to the formation of PAIs (Fig. 1):

- i. The acquisition of virulence gene(s) (often as an operon) by HGT using one of the three mechanisms described above. The source of the acquired genes is a diverse flexible gene pool, available in the environment (see below).
- ii. Following successful uptake of a foreign genetic element, an integration into the chromosome (or into an already existing plasmid) takes place, by means of site-specific recombination or other mechanisms. Integrations of plasmids and bacteriophages into the chromosome are well known phenomena (Ott, 1993), and integration of PAIs presumably occurs in a similar manner. The integration event is, probably, mediated by an integrase or recombinase, as discussed above. DNA fragments of different sizes can thus find their way into a new host and be incorporated or recombined into the core genome creating new distinct genetic island structures. It is noteworthy that some PAIs represent mosaic-like structures, rather than single homogeneous segments of horizontally acquired DNA, suggesting multiple acquisitions from different donors during evolution into the same locus.
- iii. Subsequently, a mobile genetic element might develop into a PAI by genetic rearrangements, gene loss, or further acquisition of other genetic elements. PAIs might become immobilized due to inactivation or deletion of origins of a plasmid replication (*ori*) or of genes that are involved in the mobilization and/or self-transfer of plasmids (*tra*, *mob*) or bacteriophages (*int*).



**Fig. 1.** Different stages in the evolution of a PAI. A genetic element is acquired by a bacterial cell, through a horizontal gene transfer (transduction in this case) from an environmental flexible gene pool (1). Following a successful uptake (2), recombination which, is mediated by an integrase (*int*) or other mechanisms (3) results in integration of the acquired element into the chromosome (4). Loss of the genes that are involved in the mobility of the element leads to stably integrated PAI in the chromosome (5). If the incorporated PAI confers an advantage to the organism in the specific niche, a positive selection will operate to favour this variant. Subsequently, the frequency of this variant will increase in the population over time (6). Genetic rearrangements or new gene acquisitions will lead to further evolution of the PAI. Now the modified PAI can be recombined with the environmental available gene pool and be further transferred to a new recipient microorganism (7). (The bacteriophage image was adopted and modified from <http://www.swbic.org/products/clipart/images/bacteriophage.jpg>)

The 'trade-off' of this process is a stable association and inheritance with the host chromosome.

- iv. Under certain circumstances, the newly acquired assimilated genes might be successfully expressed. If this expression contributes to the inclusive fitness of the bacteria (for instance, increase in pathogenic potency or transmission to new host); a positive selection will favour these variants. This kind of natural selection results in clonal expansion and will lead to the increase in frequency of the variants harbouring the beneficial genes in the population over time. At this stage, it is also expected that a distinct regulation of the virulence genes will evolve. Appropriate regulation, which may include newly PAI-encoded or already-existing regulators, will provide accurate expression

that will be coupled to other virulence functions of the pathogen.

- v. PAIs might further evolve by consecutive recombination, insertion, or excision events that result in gains or losses of genetic information. In this way, features of mobile genetic elements could also be regained, resulting in chromosomal excision of the entire island and enabling its transfer to another recipient.

### Origin of PAIs

A recent analysis of the genome sequences of 63 prokaryotes suggests a distinct gene pool associated with GEIs in

comparison to the core DNA genome, which is primarily vertically inherited. This analysis further showed that unique genes that have no homologues in other genomes are more likely to be present in the horizontally acquired GEIs (Hsiao *et al.*, 2005).

An environmental horizontal gene pool is most likely available in ecological niches that are colonized by diverse bacterial species, such as aquatic systems, soil, mixed microbial biofilms, or the rumens and guts of animals. These kinds of niches continuously provide their residents with 'alien' genetic substrates of 'naked' DNA segments in the form of plasmids, chromosomal DNA fragments, phages, integrating conjugative elements, and other mobile genetic elements, as well as an assortment of other genes that may move in tandem with these elements.

Nonetheless, it is interesting to note that in some cases the origin of horizontally acquired genes does not necessarily have to be from other prokaryotes. Species of *Xanthomonas* may have acquired PNP genes, encoding plant natriuretic peptides, from their host plant genomes, potentially enabling molecular mimicry (Nembaware *et al.*, 2004), and *Legionella* spp. have numerous eukaryotic-like proteins that interact with host cell proteins and have presumably derived from a eukaryotic origin (Cazalet *et al.*, 2004).

While transformation, conjugation and transduction all have been implicated as mechanisms for HGT, recent analyses have indicated that phage transduction is the predominant force in cross-taxa transfer (Canchaya *et al.*, 2003). With phage diversity approximately 10 times that of prokaryotic diversity, several researchers have proposed that phages can contribute to the genetic individuality of bacterial strains to a much higher level than previously believed. Hence, bacteriophages are now seen as a versatile carrier of new genetic information within and between bacterial species and as means of rearranging existing genetic information in unique combinations.

Bacteriophages encoding virulence factors can convert their bacterial host, through a process called 'lysogenic conversion', from a non-pathogenic strain to a virulent strain, or a strain with increased virulence. These bacteriophages harbour genes that can provide the bacterial host with diverse repertoire of proteins such as extracellular toxins, type III secretion effectors, enzymes required for intracellular survival, adhesins for bacterial host attachment, factors involved in avoiding host immune defences, and others (Boyd and Brüssow, 2002).

This notion suggests that transduction is a central player in the formation of PAIs. Examples supporting this concept are found in the *V. cholerae* VPI-1 PAI, which can be transmitted among strains by the transducing vibriophage, CT-T1 (Karaolis *et al.*, 1999). General transduction

also plays a role for transfer of SaPI1 and related islands in *Staphylococcus aureus* (Lindsay *et al.*, 1998).

### Identification of PAIs

The availability of genome data and the distinct properties of PAIs provide efficient approaches for direct identification of PAIs. Sequencing of entire bacterial genomes has permitted deeper insight into the structure and properties of bacterial chromosome and detection of horizontally acquired genetic information. Several approaches have been used to computationally identify genomic islands in sequenced genomes. One common methodology involves the identification of genomic regions that contain atypical sequence composition in comparison to the genome-wide mean, such as abnormal G + C content, dinucleotide bias, and distinct codon usage. Although this approach is relatively straightforward and simple, it may fail to detect more ancient horizontal transfer events due to the amelioration of the acquired sequences over time, or regions that were acquired from organisms with similar sequence compositions.

A different approach that can be used is to search for genes with functions that are often associated with HGT events such as mobility genes, integrases, transposases, phage genes, or genes with unusual similarity to phylogenetically distant species. tRNA genes screening also proved successful and was used to identify genomic regions specific to *Salmonella enterica* serovars Typhimurium and Typhi (Hansen-Wester and Hensel, 2002) and more recently in four *E. coli* and *Shigella* strains (Ou *et al.*, 2006). A third approach would be a direct comparative genomics of evolutionary related species or serovars and the identification of unique regions.

Several tools have been developed to facilitate *in silico* detection of (potential) PAIs ([http://gchelpdesk.ualberta.ca/news/30jan06/cbhd\\_news\\_30jan06.php#IslandPath](http://gchelpdesk.ualberta.ca/news/30jan06/cbhd_news_30jan06.php#IslandPath)). These tools include programs like 'IslandPath', an application that integrates multiple features of PAIs such as atypical sequence composition and HGT associated genes for island detection (Hsiao *et al.*, 2003).

As opposed to *in silico* analyses, other laboratory 'wet' approaches do not require preliminary sequence information. These approaches include genome comparison by subtractive hybridization, which allows the identification of strain-specific DNA regions, or assessing the stability of a DNA region using counterselectable markers. This approach was applied to *Shigella flexneri* and resulted in the identification of the SHI1 (initially defined as *she*) PAI (Rajakumar *et al.*, 1997).

Identification of distinct PAIs can have medical and practical implications as well, as genes located on PAIs can be used as markers in molecular diagnosis of bacte-

rial pathogens, estimation of their pathogenic potential, and even their antibiotic resistance pattern.

### Distribution of PAIs

An increasing number of sequenced bacterial genomes and the use of various comparative genomic approaches have revealed that PAIs are found in many phylogenetically unrelated microorganisms. The list of currently characterized PAIs is extensive. We do not attempt to list the complete inventory, but more details can be found in recent reviews (Schmidt and Hensel, 2004; Hochhut *et al.*, 2005). It is clear that PAIs are ubiquitous and can be found in both Gram-negative and Gram-positive, human, animal and plant pathogens including:

- i. Gram-negative pathogens: *H. pylori*, different *E. coli* lineages, *Salmonella* spp., *Shigella* spp., *Yersinia* spp., *Citrobacter rodentium*, *L. pneumophila*, *Pseudomonas aeruginosa*, *Pseudomonas syringae*, *V. cholerae*, *D. nodosus*, *Erwinia amylovora*, *Bacteroides fragilis* and *Porphyromonas gingivalis*.
- ii. Gram-positive pathogens: *Listeria* spp., *S. aureus*, *Streptococcus* spp., *Enterococcus faecalis* and *Clostridium difficile*.

It is noteworthy that lateral gene transfer is not limited to prokaryotes, and examples of horizontally acquired genes can be found among unicellular eukaryotes as well (Andersson, 2005). Examples include the HGT from the rumen bacterium *Fibrobacter succinogenes* to the rumen fungus *Orpinomyces joyonii* (Garcia-Vallve *et al.*, 2000). HGT has also been identified in the pathogenic fungus *Ustilago hordei* (Lee *et al.*, 1999), and in parasitic lineages of protozoa (Richards *et al.*, 2003), suggesting that further analysis of eukaryotic genomes will define more clearly the roles of these elements in the evolution of eukaryotic pathogens.

In general, the presence of a particular PAI is specific to a certain bacterial pathogen, or even to a particular strain or serovar. For example, the PAI I<sub>536</sub> and the PAI III<sub>536</sub> are strain specific to uropathogenic *E. coli* strain 536 only; whereas PAI<sub>CFT073</sub> can be found exclusively in UPEC strain CFT073 (Guyer *et al.*, 1998). On the other hand, some PAIs are rather promiscuous. A well-studied example is HPI, which was first discovered in *Yersinia* strains (Carniel *et al.*, 1996), but can be found in different pathogenic *E. coli* strains (Schubert *et al.*, 1998), as well as in *Citrobacter diversus* and different *Klebsiella* species (Bach *et al.*, 2000). Another example is the LEE PAI, which is present in different enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC) strains, as well as in *C. rodentium* (see below). It has also become obvious that certain pathogenic species or bacterial pathotype often possesses more than one PAI in its

genome. *S. enterica*, for example, carries as many as twelve characterized PAIs, and UPEC 536 and *S. flexneri* harbour at least four PAIs.

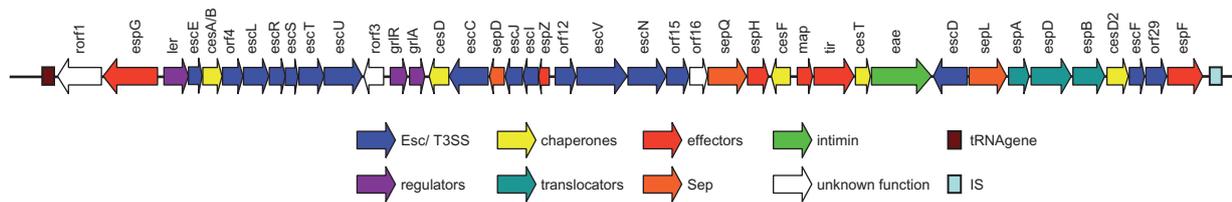
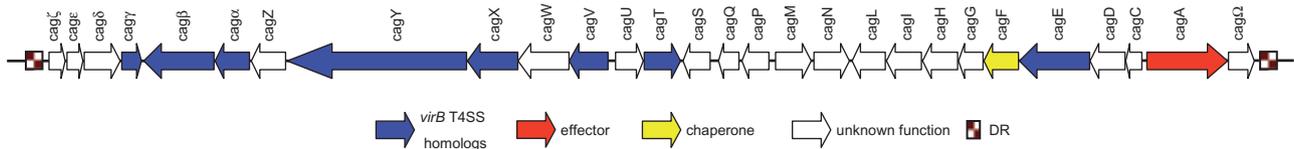
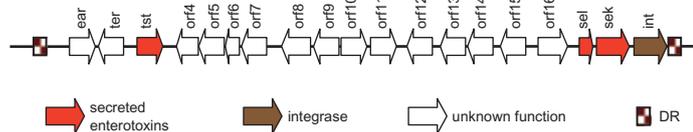
Nevertheless, PAIs appear absent in several groups of bacterial pathogens including *Mycobacterium* spp., *Chlamydia* spp. and the spirochetes. The reason for their apparent absence in certain species is not completely understood; however, examination of the lifestyle of these pathogens may give some hints about the underlying principles. The complete genome sequences of obligate intracellular pathogens (Moran, 2002) have shown that some groups of pathogens lacking PAIs are highly specialized and adapted to a specific host environment. This lifestyle is accompanied by reduction of the genome size and loss of the ability to replicate outside a host. In contrast, most pathogens harbouring PAIs show a high degree of flexibility in the utilization of different hosts or in the proliferation sites inside them. As the cost of specialization is genome reduction, it is possible that such a reduction has led to the deletion of major portions of horizontally acquired DNA elements. In addition, adaptation of a parasitic or obligate intracellular lifestyle might also result in reduced access to an environmental flexible gene pool (Schmidt and Hensel, 2004).

### Contribution to virulence

In contrast to Charles Darwin's theory, assuming that evolution of species progresses slowly by very small steps (Darwin, 1859), the concept of PAI acquisition allows 'quantum leaps' in genetic variation, and therefore in the evolution of bacterial species (Groisman and Ochman, 1996). Bacterial pathogens fitness can be defined as how well a bacterial strain can infect a host, persist, proliferate, and be transmitted to a new host in a specific niche (i.e. reproduce). In the light of this definition, an increased fitness could be achieved by a simultaneous acquisition of many genes by HGT that allow the bacteria to rapidly gain complex virulence functions and to exploit a new environmental niche (Ochman *et al.*, 2000). In some cases, introduction of a new PAI can result in a dramatic or even total change of the phenotype or lifestyle of a bacterium. The ancestor of *S. enterica*, for example, was likely an intestinal-dwelling bacterium which was not capable of invading epithelial cells. An acquisition of fully functional PAIs known as SPI 1 and subsequently SPI 2 provided *Salmonella* new physiological capabilities and was an effective strategy to make a transition towards adaptation to a new intracellular environment.

Often an acquired PAI contains an entire operon(s) that acts as a functional unit conferring new virulence traits. According to the 'selfish operon' theory (Lawrence and Roth, 1996), an ongoing selective pressure leads to the clustered organization of genes whose products contrib-

## A. EPEC LEE (35 kb)

B. *H. pylori* *cag* (37 kb)C. *S. aureus* SaPI1 (15 kb)

**Fig. 2.** The genetic organization of three representative PAIs. A schematic illustration of the LEE PAI from EPEC (A), the *cag* PAI of *H. pylori* (B), and the SaPI1 of *S. aureus* (C) is shown. The genetic nomenclature of the EPEC LEE is based on the suggested terminology by Pallen *et al.* (2005). The organization of the *cag* PAI is according to Fischer *et al.* (2001), and the organization of the SaPI1 is based on Novick (2003).

ute to a single function in order to facilitate their HGT and their propagation in the population. This kind of selection shapes gene organization and actually drives the ability of PAIs to transfer large numbers of genes in a single event.

In order to demonstrate how PAIs specifically contribute to the virulence of pathogens we examine in more detail three representative PAIs including the LEE of pathogenic *E. coli* and related species; the *cag* PAI of *H. pylori*; and the PAI encoding toxic shock syndrome toxins (TSST) of the Gram-positive pathogen *S. aureus* (SaPI).

### The LEE PAI

The locus of enterocyte effacement (LEE) was initially described in an EPEC strain, the causative agent of infant diarrhoea (McDaniel *et al.*, 1995). EPEC is an attaching and effacing (A/E) pathogen that is able to attach to host intestinal epithelium and efface brush border microvilli. All the genes necessary for this phenotype are located on a 35 kb PAI, termed LEE, which is absent from laboratory *E. coli* strains (Elliott *et al.*, 1998). Cloning the LEE into *E. coli* K-12 strain confers the complete A/E phenotype, reinforcing the notion that avirulent bacteria can be transformed into pathogenic ones through a single genetic step (McDaniel and Kaper, 1997). The LEE contains 41 ORFs (Fig. 2A) and is organized as five polycistronic operons

(LEE1–LEE5). Analysis of the G + C content of the LEE (38%) showed that it is strikingly lower than that of the rest of the chromosome (50.8%). The chromosomal integration site of LEE in the EPEC reference strain E2348/69 is the *selC* tRNA gene, but other sites are found in different EPEC strains. The LEE PAI consists of functionally different modules including: (i) a T3SS, which is used as a molecular syringe to translocate effector proteins into host cells, (ii) the secreted translocator proteins (EspA, EspD and EspB) required for translocating effectors into host cells, (iii) the adhesin (intimin, EAE), which mediates intimate attachment to Tir on the host cell cytoplasmic membrane and (iv) the secreted effector proteins EspF, EspG, EspZ, EspH, Map and Tir, the intimin receptor, chaperoned by CesT. Translocation of these molecules by T3SS into the host cells results in changes of the host cell cytoskeleton arrangement leading to the formation of actin-rich pedestals in which the Tir effector is located at their tip. This structure allows the direct interaction of Tir with the bacterial outer membrane protein intimin, as well as the host cytoskeleton (reviewed in Zaharik *et al.*, 2002).

LEE genes are controlled in a complex manner by different regulators encoded within the PAI, on a plasmid, and on the core genome. These regulators include the Ler (LEE-encoded regulator), GrIA (global regulator of LEE-

activator), GrIR (global regulator of LEE-repressor), the plasmid-encoded regulator (Per), the DNA-binding protein H-NS and the integration host factor (IHF) (Clarke *et al.*, 2003; Deng *et al.*, 2004).

Similarly to EPEC, the LEE PAI was also found in EHEC strains. The LEE PAI of EHEC encodes proteins also involved in the A/E phenotype and is inserted in the *selC* site as well. The LEE region of EHEC contains 54 ORFs, of which 41 are common with the EPEC LEE. The remaining 13 ORFs belong to a putative P4-like prophage element, designated 933 L that is located close to the *selC* locus and, probably, has been acquired at a later time point. It is interesting to note that despite EPEC and EHEC sharing virtually the same LEE PAI, their primary host, the colonization sites, and the disease they cause are different. EPEC strains are classically associated with diarrhoea in young children and humans are considered their primary host. In contrast, EHEC infections are known to originate from particular ruminants. EHEC colonization of ruminants is generally asymptomatic while in humans it can cause a spectrum of diseases ranging from uncomplicated watery diarrhoea to bloody diarrhoea with abdominal cramps. These differences likely result from the evolution of EHEC from EPEC through the acquisition of phage-encoded Shiga toxins (Stx) (Reid *et al.*, 2000).

LEE has also been identified in *C. rodentium*, the causative agent of transmissible murine colonic hyperplasia in suckling mice. Although the *C. rodentium* LEE shares 41 ORFs with EPEC and EHEC LEE, it is unique in the *rorf1* and *espG* gene location, and the presence of several ISSs. As opposed to the EHEC and EPEC LEE, *C. rodentium* LEE is not integrated into the *selC* locus and contains on one side an ABC transport system and an IS element on the other side. Based on this, it has been suggested that the LEE PAI may have been acquired several times during the evolution of different A/E pathogens.

Particular *E. coli* strains are associated with diarrhoea and other enteric infections in rabbits, pigs, calves, lambs and dogs. These EPEC strains contain LEE PAIs inserted in *selC*, *pheV* or *pheU* tRNA loci. The rabbit diarrhoeagenic *E. coli* strain RDEC-1 contains a LEE that is flanked by an IS2 element and the *lifA* toxin gene (Zhu *et al.*, 2001). LEE has also been characterized in a bovine Shiga toxin-producing *E. coli* (STEC) O103:H2 strain (Jores *et al.*, 2001).

Besides the secreted effector proteins encoded within the LEE PAI, recent work in our and others labs led to the identification of six non-LEE-encoded, conserved effector proteins (NleA, NleB, NleC, NleD, NleE and NleF). In EHEC, these effectors are organized in three additional distinct PAIs (Deng *et al.*, 2004; reviewed in Garmendia *et al.*, 2005). Recent analyses showed that *nleB* and *nleE*, encoded within a PAI known as O-Island 122, are associated with outbreaks and haemolytic–uraemic syndrome

of non-O157:H7 STEC; and that NleA and NleB are absolutely necessary to cause mortality in the mouse model (Wickham *et al.*, 2006). These observations indicate that diseases mediated by A/E pathogens require co-ordinated and regulated action of effectors encoded by the LEE and other PAIs (Deng *et al.*, 2004).

#### The *cag* PAI of *H. pylori*

Since its isolation from human stomach biopsies in 1983 (Marshall and Warren, 1984), *H. pylori* has been the focus of intense research. As a human gastric pathogen, *H. pylori* colonizes over half of the world's population. While many of *H. pylori* infected individuals are clinically asymptomatic, most will exhibit some degree of gastritis; approximately 10% of the infected subjects will develop more severe gastric pathologies like peptic ulcer disease, atrophic gastritis; and approximately 1% of infected individuals will develop gastric cancer. One of the defined *H. pylori* virulence factors is the *cag* PAI. Strains of *H. pylori* associated with severe gastric disease, such as peptic ulcer disease, possess the *cag* PAI, which is absent from strains isolated from patients with uncomplicated gastritis (Censini *et al.*, 1996). Similar correlation was also found in the infection model of Mongolian gerbils demonstrating that *H. pylori* strains with an intact *cag* PAI induced strong inflammation and ulceration in the stomach (Ogura *et al.*, 2000). Interestingly, studies with a mouse model have shown an association between *cag* PAI-negative *H. pylori* strains and strains that are mouse adapted, less virulent, and can better colonize mice, indicating that the *cag* PAI may become lost during colonization of animals (Philpott *et al.*, 2002).

The *cag* PAI is a 37–40 kb chromosomal region that was acquired by horizontal transfer and inserted at the distal end of the glutamate racemase gene (*glr*). *cag* has a distinct G + C content, and is flanked by DRs of 31 bp that probably function as sites for recombination and deletion of the locus. Sequence analysis of the *cag* PAI predicted 27 ORFs and an additional element which is not present in all of the *cag* positive strains (hp548/*cag*Ω; Fig. 2B). A large portion of the *cag* PAI genes encode a functional T4SS and eight of them are homologues to components of the prototype T4SS represented by the *A. tumefaciens* *virB* operon. In addition to the T4SS, the *cag* PAI encodes CagA, the only effector protein of the *H. pylori* T4SS currently known (Segal *et al.*, 1999). Studies of CagA's cellular activities reveal that CagA interacts with a large number of host proteins and has multiple effects on host signal transduction pathways, the cytoskeleton and cell junctions (for a recent review see Bourzac and Guillemin, 2005). After translocation of CagA into host cells, it becomes phosphorylated by Src kinases and is recruited to the plasma membrane, where it interacts with

a number of host proteins. The best studied of these interactions is with the SRC-homology 2 (SH2) domain-containing tyrosine phosphatase (SHP-2). Interaction of SHP-2 and CagA activates particular pathways and leads to actin polymerization, cell elongation, pedestal formation as well as growth factor-like response and abnormal proliferation of gastric epithelial cells. Besides SHP-2, other substrates which CagA interacts with include ZO-1, Grb2 and C-Met (reviewed in Naumann, 2005). In tissue culture cells and in the mouse model, it has been shown that the *cag* PAI induces expression of proinflammatory cytokines, such as interleukin-8 (IL-8), which is thought to contribute to *H. pylori*-induced inflammation in the stomach (Crabtree *et al.*, 1995; Philpott *et al.*, 2002). A recent study also showed an interaction between CagA and another *cag* protein, namely CagF, suggesting that CagF might function as a chaperone-like protein for CagA (Couturier *et al.*, 2006).

Systematic mutagenesis approaches to analyse the function of the 27 genes in the *cag* PAI identified a subset of 17 genes that are absolutely required for the translocation of CagA and a subset of 14 genes that are required for the stimulation of IL-8 synthesis in host cells (Fischer *et al.*, 2001). Although the assembly of the T4SS is not understood in full detail, these observations indicate that the majority of the *cag* PAI genes are required for the formation of a functional T4SS which is used to: (i) translocate the bacterial effector protein CagA into host cells, and (ii) induce the synthesis and secretion of chemokines, such as IL-8.

Cumulatively, these studies show a pivotal role of the *cag* PAI in the virulence of *H. pylori* and clearly demonstrate the way by which a single locus contributes to the pathogenic lifestyle of a bacterium.

#### Staphylococcus PAI encoding TSST

Chromosomal regions with the typical features of PAIs as in Gram-negative bacteria are apparently less abundant in Gram-positive pathogens, although some of the characteristics of PAIs have also been identified in these microorganisms. Genome comparative analyses between related Gram-positive bacteria have demonstrated that acquisitions of genomic islands are indeed the main source of pathogenicity and resistance profile differences and therefore play a similar role as in Gram-negative bacteria (Gill *et al.*, 2005).

*Staphylococcus aureus* is a common commensal bacterium found on human skin and respiratory tract mucosal surfaces. However, it is also a pathogen causing a range of acute and pyogenic infections, including abscesses, bacteraemia, central nervous system infections, endocarditis, osteomyelitis, pneumonia, urinary tract infections, chronic lung infections associated with cystic

fibrosis and several syndromes caused by a variety of toxins. These toxins, including haemolysins, staphylococcal exotoxins (Set) and superantigens (SAGs), are major virulence factors of *S. aureus*. Staphylococcal SAGs are a group of high molecular-weight proteins that are potent stimulatory agents for CD4<sup>+</sup> T lymphocytes. As such, they have profound effects on the immune system, leading to non-specific activation of a large proportion of T cells, resulting in the release of various cytokines. Certain *S. aureus* strains possess secreted virulence factors known as TSST that function as superantigen toxins (Bachert *et al.*, 2002). The consequences of these TSST may include high fever, rash, vomiting, diarrhoea, renal and hepatic dysfunction and desquamation.

The chromosomal *tst* gene, encoding TSST-1, is located on a series of discrete 15–20 kb chromosomal elements that are mobilized at high frequencies by certain staphylococcal phages. These elements are referred to as staphylococcal pathogenicity islands (SaPIs) and were the first clearly defined PAI characterized in Gram-positive bacteria (reviewed in Novick, 2003). The prototype of this family is the SaPI1 that was the first characterized SaPI. SaPI1 is 15.2 kb long, carries a *tst* gene, flanked by 17 bp DR sequences, and is inserted in an *att<sub>c</sub>* site close to the *tyrB* gene (Fig. 2C). The integration into the chromosome is facilitated by the presence of a functional integrase (*int*) gene encoded in the island. Remarkable features of SaPI1 are therefore its mobility and instability. The excision of SaPI1 from the chromosome and its presence as episomal DNA have been observed (Ruzin *et al.*, 2001).

In addition to SaPI1, other SaPIs, which carry *tst* and different SAGs genes, have been characterized. A PAI related to SaPI1, termed SaPIbov, was identified in a bovine isolate of *S. aureus*. SaPIbov is 15.9 kb long and is inserted at the 3' end of the GMP synthase gene (*gmps*), in an *att* integration site. SaPIbov is flanked by 74 bp DR sequences and harbours, in addition to *tst*, two other enterotoxins, encoded by *sec* and *sel* genes (Fitzgerald *et al.*, 2001). SaPI3 has been shown to contain two novel enterotoxins encoded by the *sek* and *seq*, as well as the enterotoxin B (SEB). SaPI3 is flanked as well by *att* sites and displays an overall structure similar to SaPI1; however, a *tst* gene is absent in SaPI3 (Yarwood *et al.*, 2002). Interestingly, the presence of *int* genes and *att* sites in the SaPIs suggests that they have been acquired from phage genomes.

A recent analysis of several *S. aureus* isolates has led to the identification (and renaming) of seven conserved PAI families in the *S. aureus* genome designated vSa1 (including SaPI1 and SaPI3); vSa2 (including SaPIbov); vSa3; vSa4 (including SaPI2); vSa $\alpha$ ; vSa $\beta$ ; and vSa $\gamma$  (Gill *et al.*, 2005). Besides the TSST encoded PAIs, another important PAI is the *Staphylococcus* cassette chromosome *mec* (SCC*mec*), which encodes the methicillin

resistance determinants MetI, MetR and MetA (Daum *et al.*, 2002).

In summary, various PAIs in the *S. aureus* genome carry approximately one-half of its toxins or virulence factors, and allelic variation of these genes, along with the presence or absence of individual islands, contributes to the pathogenic profile of *S. aureus* species (Gill *et al.*, 2005).

### Concluding remarks and open questions

Infectious diseases remain a significant cause of mortality and morbidity worldwide. This problem has been recently heightened with the increased resistance of microbes to antibiotics and the emergence of new pathogens. Identifying virulence factors used by these bacterial pathogens and understanding their evolution from their non-pathogenic progenitors are important both for basic science and for current medical challenges. In this review, we focused on a group of mobile genetic elements termed 'pathogenicity islands'. Despite tremendous insights that have been gained in the last few years, there are still gaps in our knowledge and many open questions have yet to be addressed.

One of these unknown aspects is the origin of some of the PAIs. The LEE PAI, for instance, is speculated to have been acquired independently several times by different pathogens (EPEC, EHEC and *Citrobacter*), suggesting that its donor was available in the environment for a long period of time. However, what was the nature of this organism, and is it still circulating in the environment or has become extinct, is not known.

Other intriguing questions are related to the regulation of the PAIs encoded genes. Newly acquired sequences may decrease fitness unless integrated into pre-existing regulatory networks. Therefore, the acquisition of PAIs, even when they encode their own specific regulators, must interact with the rest of the genome. This complex regulatory scheme may include regulators encoded within the islands, elsewhere on the chromosome or on plasmids, as was demonstrated for the LEE PAI. Like other virulence genes, PAI genes are usually expressed in response to environmental signals. Expression of six invasion genes within the SPI 1 island of *Salmonella* is controlled by the conserved PhoP/Q regulatory system (Galan, 1996), which is encoded outside of SPI 1 and present in both pathogenic and non-pathogenic bacteria. In *Shigella* spp. the invasion genes, which are encoded on the *Shigella* virulence plasmid, are regulated by a chromosomal locus encoding a histone-like protein H1. Other examples are the ToxT and YbtA transcriptional activators of *V. cholerae* and *Yersinia* spp., respectively, which are located on PAIs but regulate genes outside of these regions. These examples highlight the query of how such

regulation has evolved, and which mechanisms are involved in synchronizing newly acquired genes with the global regulatory network of the pathogen. A recent study has shed some light on these intriguing questions and showed that in *Salmonella*, the nucleoid protein H-NS selectively silences horizontally acquired genes by targeting sequences with G + C content lower than the core genome including SPI-2, SPI-3 and SPI-5 (Navarre *et al.*, 2006). Nonetheless, more research is needed in order to understand these phenomena.

A different issue, which has practical implications for medicine, is how many of the emerging bacterial diseases and antibiotic resistance strains are actually driven by the acquisition of new PAIs and what are the dynamics of this process.

Several islands such as SaPI, *cag* and PAI I and II of UPEC seem to have a tendency to delete from the chromosome. It has been speculated that the loss of virulence determinants may play a crucial role during the transition from an acute state of disease to chronic infections (Blum *et al.*, 1994; Hacker and Kaper, 2000), as was shown for the *cag* PAI of *H. pylori* (Philpott *et al.*, 2002). It will be of interest to further investigate what signals trigger such changes, the extent of this phenomenon among other pathogens or symbiotic microorganisms, the forces driving it, and what role the host plays in this process.

In conclusion, recent studies have emphasized the function of PAIs as a sophisticated and modular toolbox in bacterial pathogenesis. Understanding the concept of PAIs has deeply affected the way we perceive bacterial virulence and microbial evolution. Further studies including whole genome research and advanced bioinformatics of comparative genomic analysis will further contribute to our understanding of evolution of prokaryotes (and eukaryotes) and will augment our understanding of the disease process as a result of complex interactions between the host and its pathogen.

### Acknowledgements

We would like to thank to Dr Wanyin Deng and members of the Finlay lab for helpful discussions and insightful comments on the manuscript. The work in the Finlay lab is supported by operating grants to B.B.F. from the Canadian Institutes of Health Research (CIHR) and the Howard Hughes Medical Institute (HHMI). O.G. is a recipient of a Postdoctoral Fellowship from the Michael Smith Foundation for Health Research (MSFHR). B.B.F. is a CIHR Distinguished Investigator, an HHMI International Research Scholar and the University of British Columbia Peter Wall Distinguished Professor.

### References

- Andersson, J.O. (2005) Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* **62**: 1182–1197.

- Bach, S., de Almeida, A., and Carniel, E. (2000) The *Yersinia* high-pathogenicity island is present in different members of the family Enterobacteriaceae. *FEMS Microbiol Lett* **183**: 289–294.
- Bachert, C., Gevaert, P., and van Cauwenberge, P. (2002) *Staphylococcus aureus* enterotoxins: a key in airway disease? *Allergy* **57**: 480–487.
- Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschape, 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.
- Bourzac, K.M., and Guillemin, K. (2005) *Helicobacter pylori*–host cell interactions mediated by type IV secretion. *Cell Microbiol* **7**: 911–919.
- Boyd, E.F., and Brüssow, H. (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* **10**: 521–529.
- Buchrieser, C., Prentice, M., and Carniel, E. (1998) The 102-kilobase unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *J Bacteriol* **180**: 2321–2329.
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., and Brüssow, H. (2003) Phage as agents of lateral gene transfer. *Curr Opin Microbiol* **6**: 417–424.
- 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.
- Cazalet, C., Rusniok, C., Bruggemann, H., Zidane, N., Magnier, A., Ma, L., et al. (2004) Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* **36**: 1165–1173.
- Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., et al. (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA* **93**: 14648–14653.
- 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.
- Clarke, S.C., Haigh, R.D., Freestone, P.P., and Williams, P.H. (2003) Virulence of enteropathogenic *Escherichia coli*, a global pathogen. *Clin Microbiol Rev* **16**: 365–378.
- Couturier, M.R., Tasca, E., Montecucco, C., and Stein, M. (2006) Interaction with CagF is required for translocation of CagA into the host via the *Helicobacter pylori* type IV secretion system. *Infect Immun* **74**: 273–281.
- Crabtree, J.E., Covacci, A., Farmery, S.M., Xiang, Z., Tompkins, D.S., Perry, S., et al. (1995) *Helicobacter pylori* induced interleukin-8 expression in gastric epithelial cells is associated with CagA positive phenotype. *J Clin Pathol* **48**: 41–45.
- Darwin, C. (1859) The origin of species. In *A Darwin Selection*. Ridley, M. (ed.). London: Fontana Press, 1994, pp. 84–135.
- Daum, R.S., Ito, T., Hiramatsu, K., Hussain, F., Mongkolratthanotai, K., Jamklang, M., and Boyle-Vavra, S. (2002) A novel methicillin-resistance cassette in community-acquired methicillin-resistant *Staphylococcus aureus* isolates of diverse genetic backgrounds. *J Infect Dis* **186**: 1344–1347.
- Deng, W., Puente, J.L., Gruenheid, S., Li, Y., Vallance, B.A., Vazquez, A., et al. (2004) Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc Natl Acad Sci USA* **101**: 3597–3602.
- Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* **284**: 2124–2129.
- Elliott, S.J., Wainwright, L.A., McDaniel, T.K., Jarvis, K.G., Deng, Y.K., Lai, L.C., et al. (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* **28**: 1–4.
- 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.
- Fischer, W., Puls, J., Buhrdorf, R., Gebert, B., Odenbreit, S., and Haas, R. (2001) Systematic mutagenesis of the *Helicobacter pylori* *cag* pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* **42**: 1337–1348.
- Fitzgerald, J.R., Monday, S.R., Foster, T.J., Bohach, G.A., Hartigan, P.J., Meaney, W.J., and Smyth, C.J. (2001) Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *J Bacteriol* **183**: 63–70.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- Galan, J.E. (1996) Molecular genetic bases of *Salmonella* entry into host cells. *Mol Microbiol* **20**: 263–271.
- Garcia-Vallve, S., Romeu, A., and Palau, J. (2000) Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol Biol Evol* **17**: 352–361.
- Garmendia, J., Frankel, G., and Crepin, V.F. (2005) Enteropathogenic and enterohemorrhagic *Escherichia coli* Infections: translocation, translocation, translocation. *Infect Immun* **73**: 2573–2585.
- Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., Deboy, R.T., Ravel, J., et al. (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J Bacteriol* **187**: 2426–2438.
- Groisman, E.A., and Ochman, H. (1996) Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**: 791–794.
- Guyer, D.M., Kao, J.S., and Mobley, H.L. (1998) Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. *Infect Immun* **66**: 4411–4417.
- Hacker, J., and Carniel, E. (2001) Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* **2**: 376–381.

- Hacker, J., and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**: 641–679.
- 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 vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* **8**: 213–225.
- Hacker, J., Blum-Oehler, G., Muhldorfer, I., and Tschape, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**: 1089–1097.
- Hansen-Wester, I., and Hensel, M. (2002) Genome-based identification of chromosomal regions specific for *Salmonella* spp. *Infect Immun* **70**: 2351–2360.
- Hochhut, B., Dobrindt, U., and Hacker, J. (2005) Pathogenicity islands and their role in bacterial virulence and survival. *Contrib Microbiol* **12**: 234–254.
- Hsiao, W., Wan, I., Jones, S.J., and Brinkman, F.S. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* **19**: 418–420.
- Hsiao, W.W., Ung, K., Aeschliman, D., Bryan, J., Finlay, B.B., and Brinkman, F.S. (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet* **1**: e62.
- Jain, R., Rivera, M.C., Moore, J.E., and Lake, J.A. (2002) Horizontal gene transfer in microbial genome evolution. *Theor Popul Biol* **61**: 489–495.
- Jores, J., Rumer, L., Kiessling, S., Kaper, J.B., and Wieler, L.H. (2001) A novel locus of enterocyte effacement (LEE) pathogenicity island inserted at *pheV* in bovine Shiga toxin-producing *Escherichia coli* strain O103: H2. *FEMS Microbiol Lett* **204**: 75–79.
- Karaolis, D.K., Somara, S., Maneval, D.R., Jr, Johnson, J.A., and Kaper, J.B. (1999) A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**: 375–379.
- Lawrence, J.G. (2005) Horizontal and vertical gene transfer: the life history of pathogens. *Contrib Microbiol* **12**: 255–271.
- Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**: 1843–1860.
- Lee, N., Bakkeren, G., Wong, K., Sherwood, J.E., and Kronstad, J.W. (1999) The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. *Proc Natl Acad Sci USA* **96**: 15026–15031.
- Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., and Novick, R.P. (1998) The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol Microbiol* **29**: 527–543.
- McDaniel, T.K., and Kaper, J.B. (1997) A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol Microbiol* **23**: 399–407.
- 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.
- Marshall, B.J., and Warren, J.R. (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **1**: 1311–1315.
- Moran, N.A. (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583–586.
- Naumann, M. (2005) Pathogenicity island-dependent effects of *Helicobacter pylori* on intracellular signal transduction in epithelial cells. *Int J Med Microbiol* **295**: 335–341.
- Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., and Fang, F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **313**: 236–238.
- Nembaware, V., Seoighe, C., Sayed, M., and Gehring, C. (2004) A plant natriuretic peptide-like gene in the bacterial pathogen *Xanthomonas axonopodis* may induce hyperhydration in the plant host: a hypothesis of molecular mimicry. *BMC Evol Biol* **4**: 10.
- Novick, R.P. (2003) Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid* **49**: 93–105.
- Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
- Odenbreit, S., and Haas, R. (2002) *Helicobacter pylori*: impact of gene transfer and the role of the *cag* pathogenicity island for host adaptation and virulence. *Curr Top Microbiol Immunol* **264**: 1–22.
- Ogura, K., Maeda, S., Nakao, M., Watanabe, T., Tada, M., Kyutoku, T., *et al.* (2000) Virulence factors of *Helicobacter pylori* responsible for gastric diseases in Mongolian gerbil. *J Exp Med* **192**: 1601–1610.
- Ott, M. (1993) Dynamics of the bacterial genome: deletions and integrations as mechanisms of bacterial virulence modulation. *Zentralbl Bakteriol* **278**: 457–468.
- Ou, H.Y., Chen, L.L., Lonnen, J., Chaudhuri, R.R., Thani, A.B., Smith, R., *et al.* (2006) A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res* **34**: 1e3.
- Pallen, M.J., Beatson, S.A., and Bailey, C.M. (2005) Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion. *BMC Microbiol* **5**: 9.
- Philpott, D.J., Belaid, D., Troubadour, P., Thiberge, J.M., Tankovic, J., Labigne, A., and Ferrero, R.L. (2002) Reduced activation of inflammatory responses in host cells by mouse-adapted *Helicobacter pylori* isolates. *Cell Microbiol* **4**: 285–296.
- Rajakumar, K., Sasakawa, C., and Adler, B. (1997) Use of a novel approach, termed island probing, identifies the *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect Immun* **65**: 4606–4614.
- Reid, S.D., Herbelin, C.J., Bumbaugh, A.C., Selander, R.K., and Whittam, T.S. (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**: 64–67.
- Reiter, W.D., Palm, P., and Yeats, S. (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* **17**: 1907–1914.
- Richards, T.A., Hirt, R.P., Williams, B.A., and Embley, T.M. (2003) Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**: 17–32.

- Ruzin, A., Lindsay, J., and Novick, R.P. (2001) Molecular genetics of SaPI1-a mobile pathogenicity island in *Staphylococcus aureus*. *Mol Microbiol* **41**: 365–377.
- Schmidt, H., and Hensel, M. (2004) Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* **17**: 14–56.
- Schubert, S., Rakin, A., Karch, H., Carniel, E., and Heesemann, J. (1998) Prevalence of the 'high-pathogenicity island' of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect Immun* **66**: 480–485.
- Segal, E.D., Cha, J., Lo, J., Falkow, S., and Tompkins, L.S. (1999) Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc Natl Acad Sci USA* **96**: 14559–14564.
- Welch, R.A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **99**: 17020–17024.
- Wickham, E.M., Lupp, C., Mascarenhas, M., Vázquez, A., Coombes, K.B., Brown, F.N., *et al.* (2006) Bacterial genetic determinants of non-O157 STEC outbreaks and hemolytic-uremic syndrome after infection. *J Infect Dis* (in press).
- Yarwood, J.M., McCormick, J.K., Paustian, M.L., Orwin, P.M., Kapur, V., and Schlievert, P.M. (2002) Characterization and expression analysis of *Staphylococcus aureus* pathogenicity island 3. Implications for the evolution of staphylococcal pathogenicity islands. *J Biol Chem* **277**: 13138–13147.
- Zaharik, M.L., Gruenheid, S., Perrin, A.J., and Finlay, B.B. (2002) Delivery of dangerous goods: type III secretion in enteric pathogens. *Int J Med Microbiol* **291**: 593–603.
- Zhu, C., Agin, T.S., Elliott, S.J., Johnson, L.A., Thate, T.E., Kaper, J.B., and Boedeker, E.C. (2001) Complete nucleotide sequence and analysis of the locus of enterocyte Effacement from rabbit diarrheagenic *Escherichia coli* RDEC-1. *Infect Immun* **69**: 2107–2115.