The true extent of prokaryote diversity, encompassing the spectrum of variability among bacteria, remains unknown. Early discussions on prokaryote diversity were frequently devoted to sterile arguments about 'how much?' or 'how many?'. Increasingly, however, the focus is turning towards trying to understand why prokaryote diversification occurs, its underlying mechanisms, and its likely impact. The significance of such studies has a broad appeal, and the popular scientific press frequently highlights such topics as the emergence of new diseases, the attribution of existing diseases to hitherto unrealized actions of prokaryotes, and the activities of prokaryotes in key environmental processes. The dynamic nature of the prokaryotic world, and continuing advances in the technological tools available to this field of study, ensure that the latest story illustrating prokaryote diversity is never far away. This book will appeal to a wide variety of microbiologists. Its coverage ranges from studies of prokaryotes in specialized environmental niches to broad examinations of prokaryote evolution and diversity, and the mechanisms underlying them. Topics include: bacteria of the gastrointestinal tract, unculturable organisms in the mouth and in the soil, the question of a link between chlamydia and heart disease, organisms from extreme environments, the diversity of archaea and their phages, comparative genomics and the emergence of pathogens, spread of genomic islands between clinical and environmental organisms, core genes, minimal genomes needed for life, horizontal gene transfer, genomic islands and the evolution of catabolic pathways, phenotypic innovation, and patterns and extent of biodiversity.

Niall A. Logan is Professor of Systematic Bacteriology in the School of Life Sciences, Glasgow Caledonian University, UK.

Hilary M. Lappin-Scott is Professor of Environmental Microbiology in the School of Biosciences, University of Exeter, UK.

Petra C. F. Oyston is Team Leader in Molecular Bacteriology at Dstl Porton Down, Salisbury, UK.

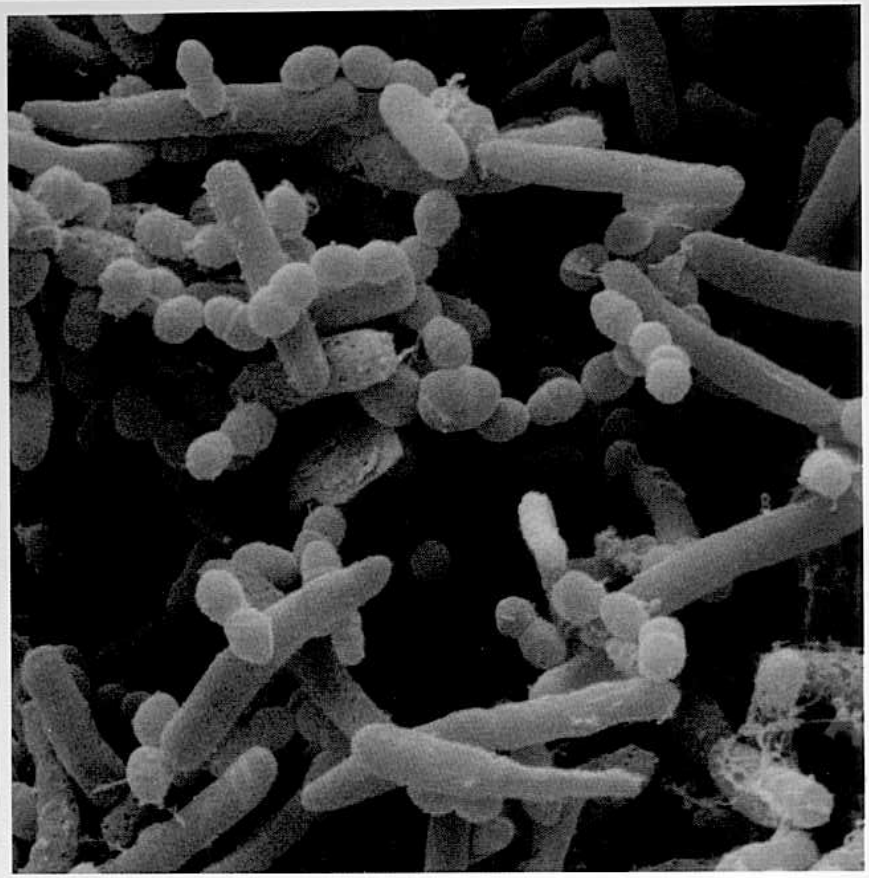
CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 0-521-86935-8

prokaryotic diversity: mechanisms and significance

CAMBRIDGE

Edited by
N. A. Logan, H. M. Lappin-Scott & P. C. F. Oyston



prokaryotic diversity: mechanisms and significance

clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O : 1b. *Mol Microbiol* **37**, 316–330.

Zhou, D. S., Han, Y. P., Song, Y. J., Huang, P. T. & Yang, R. F. (2004). Comparative and evolutionary genomics of *Yersinia pestis*. *Microbes Infect* **6**, 1226–1234.

Spread of genomic islands between clinical and environmental strains

Jens Klockgether, Oleg N. Reva and Burkhard Tümmler

Klinische Forschergruppe, OE 6711, Medizinische Hochschule Hannover, D-30625 Hannover, Germany

COMMON FEATURES OF GENOMIC ISLANDS

The genome of a bacterium consists of a core that is common to all strains of a taxon and an accessory part that varies within and among clones of a taxon. The accessory genome represents the flexible gene pool that frequently undergoes acquisition and loss of genetic information and hence plays an important role for the adaptive evolution of bacteria (Dobrindt *et al.*, 2004). The flexible gene pool is made up of accessory elements such as bacteriophages, plasmids, IS elements, transposons, conjugative transposons, integrons and genomic islands (GEIs).

GEIs are chromosomal regions that are typically flanked by direct repeats and inserted at the 3′ end of a tRNA gene. They contain transposase or integrase genes that are required for chromosomal integration and excision and further mobility-related genes. GEIs are clone- or strain-specific and are never found in all clones of a taxon. Most GEIs are easily differentiated from the core genome by their atypical G+C contents and atypical oligonucleotide composition, with steep gradients thereof at their boundaries (Reva & Tümmler, 2005). First identified in pathogenic bacteria ('pathogenicity islands'), GEIs have since been detected in numerous non-pathogenic species. GEIs may confer fitness traits, increase metabolic versatility or adaptability or promote bacterium–host interaction in terms of symbiosis, commensalism or virulence (Dobrindt *et al.*, 2004).

GEIs have been found in the majority of all currently completely sequenced bacterial

communities. Colonization of habitats with large and complex bacterial populations such as the rhizosphere or the gastrointestinal tract of animals seems to be associated with a higher frequency of GEIs than life in an isolated and/or sparsely populated niche. In other words, access to the gene pool of other taxa facilitates the emergence of GEIs.

THE PARADIGM: GENOMIC ISLANDS IN THE *ENTEROBACTERIACEAE*

Acquisition of GEIs by horizontal gene transfer could be a key factor for the adaptation of a bacterium to a particular host or niche. This issue has been addressed for the various pathovars of *Escherichia coli* (Dobrindt *et al.*, 2003; Welch *et al.*, 2002). For uropathogenic strains of *E. coli*, island acquisition has resulted in the capability to infect the urinary tract and bloodstream and evade host defences without compromising the ability for harmless colonization of the intestine. For different intestinal pathogens, acquired genes promote the colonization of specific regions of the intestine and new modes of interaction with the host tissue that produce clinically distinct variations of gastrointestinal disease. Each type of *E. coli* possesses combinations of island genes that confer its characteristic lifestyle or disease-causing traits. Examination of extraintestinal pathogenic, intestinal pathogenic and commensal *E. coli* isolates revealed a weak association between the tropism for a niche and the repertoire of pathogenicity islands (Dobrindt *et al.*, 2003). However, even though similar virulence genes come into play, their linkage relationships and chromosomal locations vary considerably from strain to strain within a pathovar (Welch *et al.*, 2002). In other words, two GEIs inserted at the same tRNA site do not necessarily contain the same set of genes.

A further example of the role of GEIs for adaptation to a host is the evolution of *Salmonella* serovars (Kingsley *et al.*, 2000). The acquisition of an island containing the *shdA* gene by the major evolutionary lineage within the genus *Salmonella*, *Salmonella enterica* subspecies I, was accompanied by an expansion in host range to include the warm-blooded mammals and birds in addition to reptiles.

Most GEIs have only been described for the sequenced index case. An interesting exception is the high-pathogenicity island (HPI) of enterobacteria. This 35–45 kb island carries genes involved in synthesis, regulation and transport of the siderophore yersiniabactin. Initially detected in pathogenic *Yersinia* species (Carniel *et al.*, 1996), the HPI was found to be harboured with high frequency by *S. enterica* subspecies III and VI (Oelschlaeger *et al.*, 2003) and enteroaggregative and extraintestinal pathogenic *E. coli* (Schubert *et al.*, 1998). However, the HPI was also detected in non-pathogenic and commensal members of the family *Enterobacteriaceae* such as *Citrobacter diversus* and *Klebsiella* species (Bach *et al.*, 2000). Iron is essential for almost all bacteria and, consequently, siderophore production has a dual role to increase fitness in iron-restricted environments and to contribute to virulence in infected hosts, which explains the widespread distribution of the HPI in pathogenic and non-pathogenic members of the *Enterobacteriaceae*.

Most HPIs are stably integrated into the chromosome, probably because of deletion or mutations in essential mobility genes. However, an HPI was recently identified in the *E. coli* strain ECOR31 that carries a complete set of conjugative plasmid functions (Schubert *et al.*, 2004). HPI_{ECOR31} was found to be inserted into a tRNA gene and carries an intact *attR* site. The GEI carries an integrase gene, an origin of transfer and DNA-processing region and a complete and functional mating-pair formation system. Induction of the HPI integrase results in precise excision and circularization of HPI_{ECOR31}. Replication as an autonomous plasmid is not possible, however, because HPI_{ECOR31} lacks the essential *repABC* genes. In summary, HPI_{ECOR31} may be considered as the progenitor of the contemporary HPIs, most of which have lost the ability to be excised because of the loss of the *attR* site, truncation of the GEI and/or deletions within the integrase gene (Schubert *et al.*, 2004).

GENOMIC ISLANDS IN *BURKHOLDERIA* AND *PSEUDOMONAS*

GEIs shared by clinical and environmental isolates

The occurrence of a GEI in pathogenic and non-pathogenic isolates has been observed not only for the HPI, but also for genome islands that are prevalent in *Pseudomonas* and *Burkholderia* species. *Burkholderia pseudomallei* is the causative agent of melioidosis. The genome of strain K96243 is composed of two chromosomes (Holden *et al.*, 2004). Sixteen GEIs together make up 6·1 % of the genome, only one of which is also present in the genome of the clonally related organism *Burkholderia mallei*. Further analysis revealed these islands to be variably present in a collection of clinical and soil isolates (Holden *et al.*, 2004). One GEI was present in all tested strains, four GEIs were detected in 10–70 % of strains, three islands were in 10 % or fewer of strains and one island was unique to the sequenced strain.

The ubiquitous and metabolically versatile *Pseudomonas aeruginosa* is an important opportunistic pathogen for humans, plants and animals. Several large GEIs have been detected in strains from human infections and aquatic habitats. The only large genome island known so far that is not associated with a tRNA gene is the 49 kb PAGI-1. PAGI-1 contains genes potentially involved in oxidative stress resistance. This first described GEI in *P. aeruginosa* was found to be present in 85 % of tested North

American clinical isolates from septicaemia, airways and urinary tract infections (Liang *et al.*, 2001). The island was probably assembled from two ancestral components of different G+C content. Thirty-five kb of the higher G+C content portion is also found in the genome of the biosafety strain *Pseudomonas putida* KT2440 (Nelson *et al.*, 2002). *P. putida* is a metabolically versatile, rapidly growing bacterium frequently isolated from moist temperate soils and waters, particularly polluted soils. The sequenced KT2440 strain is a plasmid-cured derivative of the Japanese soil isolate mt-2 (Nakazawa, 2002). In other words, PAGI-1 was identified in pseudomonads of unrelated habitats and geographical origin.

The other known large GEIs of *P. aeruginosa* integrate into tRNA genes. Two different types have been identified, the prototypes of which are the islands PAGI-2/PAGI-3 (Larbig *et al.*, 2002) and pKLC102 (Klockgether *et al.*, 2004)/PAPI-1 (He *et al.*, 2004), respectively. PAGI-2 and PAGI-3 were sequenced in two strains of the major clone C (Römling *et al.*, 1997), an isolate from the lungs of a patient with cystic fibrosis and an isolate from a river. In both strains, the region consists of an individual strain-specific GEI and a shorter stretch of clone-specific sequence. The left boundary of the islands is a cluster of tRNA genes comprising one $tRNA^{Glu}$ gene followed by two identical $tRNA^{Gly}$ genes, one serving as the integration site for the *P. aeruginosa* genome island PAGI-2, the other for PAGI-3. PAGI-2 and PAGI-3 terminate at the right end with the terminal 16 and 24 nucleotides of the 3′ end of the $tRNA^{Gly}$ gene, respectively. In both islands, the first open reading frame (ORF) adjacent to the $tRNA^{Gly}$ gene encodes a bacteriophage P4-related multidomain integrase with an unusual transposase-like C terminus. PAGI-2 and PAGI-3 have a modular bipartite structure. The first part adjacent to the tRNA gene consists of strain-specific ORFs encoding metabolic functions and transporters, the majority of which has homologues of known function in other eubacteria (ORFs C2–C35; cargo region). The second part is made up of a syntenic set of ORFs the majority of which are either classified as conserved hypotheticals or related to DNA replication or mobility genes (ORFs C1, C36–C111; conserved part). Forty-seven of these ORFs are arranged in the same order in the two islands, with an amino acid sequence identity of 35–88 %.

The incidence of islands of PAGI-2 type in the *P. aeruginosa* population has been investigated by Southern analysis (J. Klockgether, unpublished data). An array of PAGI-2 ORFs was probed with genomic DNA from 71 *P. aeruginosa* strains, each representing a separate clone (Morales *et al.*, 2004). The collection consisted of 55 clinical isolates and 16 environmental isolates of diverse geographical origin. PAGI-2-like islands were identified in 31 of the 71 strains. The ORFs of the cargo region were absent in all but five strains. ORFs of the conserved part, however, were present to different extents. Twenty-five ORFs were present in all PAGI-2-type islands, a further 23 ORFs were detected in the majority of strains and 19 ORFs were variably present. The copy number of PAGI-2-type islands in the chromosome was one (12 strains), two (11 strains), three (seven strains) or four (one strain). PAGI-2 was completely conserved in five analysed *P. aeruginosa* strains, an isolate from an ear infection from the USA isolated in 1980, an isolate from 1985 from the airways of a patient with cystic fibrosis living in Lower Saxony, Germany, and three isolates from river and sanitary facilities in households sampled in Northrhine-Westphalia, Germany, in 1992. In other words, PAGI-2 was identified in unrelated strains from diverse habitats and geographical origin.

The *P. aeruginosa* chromosome consists of three large hypervariable regions (Römling *et al.*, 1995, 1997), one of which is the PAGI-2-like GEI. The integration and excision of GEIs into the two copies of a $tRNA^{Lys}$ gene make up the other two hypervariable regions. The sequenced prototypes are the pathogenicity island PAPI-1 (He *et al.*, 2004) and the mobile genetic element pKLC102 (Klockgether *et al.*, 2004). PAPI-1 and pKLC102 share numerous features: approximate size (108 and 102 kb), a $tRNA^{Asp}$, $tRNA^{Pro}$ and $tRNA^{Lys}$ gene cluster at their leftward PAO1 junction and a direct repeat of the 3′ half of the $tRNA^{Lys}$ gene at their right border and the presence of the integrase and chromosome-partitioning genes at the ends of the island, similar to PAGI-2 and PAGI-3. PAPI-1 encodes at least 19 virulence factors that occur on GEIs found in a wide spectrum of other pathogenic bacteria (He *et al.*, 2004). PAPI-1 was first identified in wound and cystic fibrosis isolates from the USA (He *et al.*, 2004), but meanwhile has also been detected in clinical and wastewater isolates from Europe (L. Wiehlmann, personal communication). pKLC102 is more abundant than PAPI-1 in the *P. aeruginosa* population. More than 85 % of strains in our reference collection harbour pKLC102-type GEIs. Clone C strains from the environment were found to harbour chromosomal and episomal copies of pKLC102. pKLC102 contains the 8·5 kb *chvB* gene, homologues of which are known to confer host tropism and virulence and to be essential for the interaction of the bacterium with its eukaryotic host. PAPI-1 and pKLC102 encode type IV group B pili and type IV thin sex pili, respectively, and share a set of homologues found as island-specific genes in PAGI-2 and PAGI-3. The mobile pKLC102 shares with PAPI-1 the phage module that confers integrase, the *att* element and the syntenic set of conserved genes that was first detected in PAGI-2 and PAGI-3, but it differs from PAPI-1 in carrying a plasmid module that confers *oriV* and genes for replication, partitioning and conjugation (Klockgether *et al.*, 2004). In summary, PAPI-1 and pKLC102 can be found in *P. aeruginosa* isolates from both clinical and environmental habitats. PAPI-1 is associated with one abundant epidemic clone represented by the sequenced strain PA14, whereas the mobile pKLC102 is widespread in the *P. aeruginosa* population. Moreover, pKLC102-like GEIs were detected in two sequenced *Pseudomonas syringae* strains (Feil *et al.*, 2005).

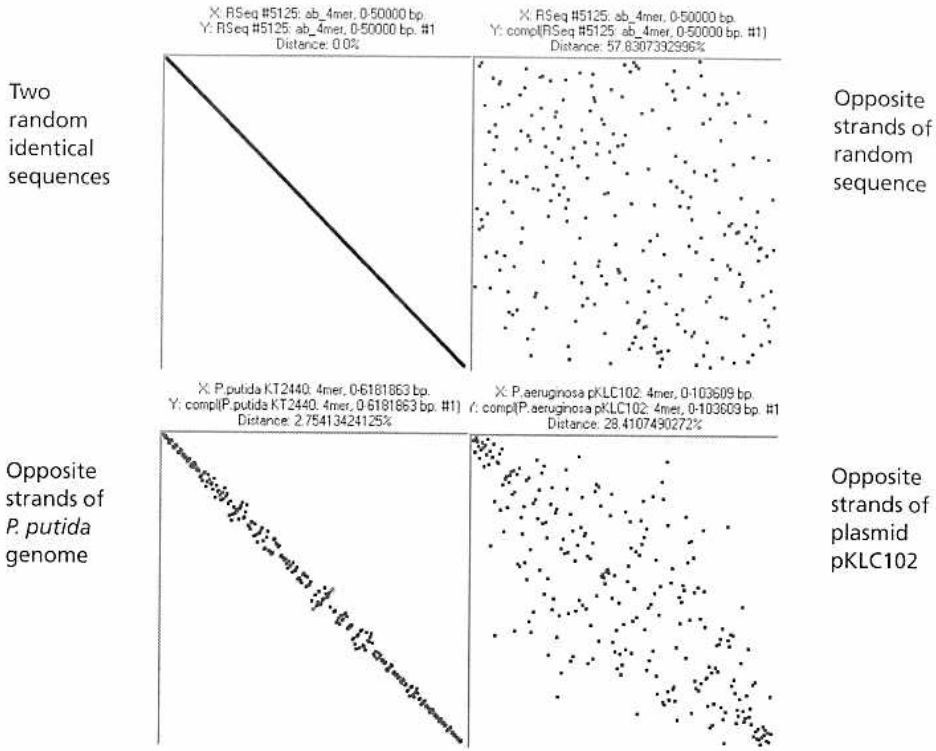


Fig. 1. Frequency of tetranucleotides in two identical sequences (50 kb, upper left) and the two complementary strands of a random sequence (50 kb, upper right), the *P. putida* KT2440 chromosome (6182 kbp, lower left) and the conjugative GEI pKLC102 (104 kbp, lower right). The frequency of each of the 256 tetranucleotides in each strand was counted and then the tetranucleotides were sorted by increasing frequency. The global difference between the oligonucleotide patterns of two strands is described by the distance *D*, i.e. the ratio of the observed to the maximal possible difference between the oligonucleotide patterns. *D* was calculated to be 0 for the identical sequence, 57·8 % for the random sequence, 2·8 % for the KT2440 chromosome and 28·4 % for pKLC102. Definition of *D* and algorithms are described in detail in the original publication by Reva & Tümmler (2004).

The promiscuous pKLC102 is a conjugative GEI of plasmid and phage origin. Compared with the typical GEI, it is endowed with exceptionally high mobility. Inspection of its oligonucleotide composition provides a hint of why pKLC102 has not been stably captured by the host chromosome. Bacterial chromosomes are characterized by strand symmetry and intrastrand parity of complementary oligonucleotides (Reva & Tümmler, 2004). In other words, oligonucleotides and their reverse complements occur with similar frequency in bacterial chromosomes. In contrast, no such correlation is observed for a random sequence (Fig. 1). Conjugative GEIs such as pKLC102 are intermediate between chromosome and random sequence. Oligonucleotide frequency on the two strands is only weakly correlated in pKLC102, whereas each oligonucleotide occurs with almost identical frequency on the two strands of a bacterial chromosome, as shown in Fig. 1 for *P. putida* KT2440. Stably integrated GEIs have an atypical oligonucleotide composition compared with the core genome, but strand symmetry is locally maintained (Reva & Tümmler, 2005). Conjugative GEIs, like most phages (Reva & Tümmler, 2004), do not adhere to this rule, suggesting that they behave as selfish parasitic DNA that is prone to horizontal spread within and among taxa.

PAGI-2 and pKLC102 belong to a family of ancient GEIs

PAGI-2, PAGI-3, PAPI-1 and pKLC102 share a syntenic set of homologous genes the majority of which are conserved hypotheticals of unknown function. A database search revealed sequence similarity to 30 further contiguous sequences in beta- and gamma-proteobacteria (Table 1; updated 10 October 2005). All GEIs were integrated into tRNA genes. Fifteen of 33 core genes were congruent, with the phylogenetic relationships of each of the individual genes indicating that all five large genome islands known so far in *P. aeruginosa* belong to one family of related syntenic GEIs with a deep evolutionary origin (Mohd-Zain *et al.*, 2004). As an example of a conserved core gene, Fig. 2 shows the phylogenetic relationships of all homologues of *P. aeruginosa* C ORF47. The core genes are likely to account for the conjugative transfer of the GEIs and may even encode autonomous replication.

Accessory gene clusters are nestled among the core genes and encode the following diverse major attributes: antibiotic, metal and antiseptic resistance, degradation of chemicals, type IV secretion systems, two-component signalling systems, Vi antigen capsule synthesis, toxin production and a wide range of metabolic functions. Examples are the large conjugative *Haemophilus influenzae* antibiotic-resistance element, ICE*Hin1056* (Mohd-Zain *et al.*, 2004), the pathogenicity island SPI-7 found in *S. enterica* serovar Typhi C18 and Ty2 (Pickard *et al.*, 2003) and the *clc* element that was originally discovered in *Pseudomonas* sp. B13 and encodes the enzymes for 3- and 4-chlorocatechol degradation (Frantz & Chakrabarty, 1987). The $tRNA^{Gly}$-integrated *clc* element is an example of a xenobiotic-degradation island that allows bacteria to handle compounds that have been recently released into the environment by humans. This type of element has been detected in numerous chloroaromatic-degrading bacteria (van der Meer & Sentchilo, 2003) including the completely sequenced *Burkholderia xenovorans* strain LB400, a micro-organism known for its capabilities to degrade polychlorobiphenyls (Bopp, 1986).

Spread of GEIs among species

Spread of GEIs among taxa has been detected for HPI (see above), the *clc* element (Ravatn *et al.*, 1998; Springael *et al.*, 2002), pKLC102 (Klockgether *et al.*, 2004; Feil *et al.*, 2005) and PAGI-2 (J. Klockgether, unpublished data). PAGI-2 and the *clc* element of *Pseudomonas* sp. B13 share 85–100 % nucleotide sequence identity in the conserved

Table 1. pKLC102-type GEIs in proteobacteria

Strain	GEI	Source/habitat	GenBank accession no.
Azoarcus sp. EbN1*		Anoxic freshwater mud	NC_006513
Azotobacter vinelandii AvOP		Soil	NZ_AAAU00000000
Burkholderia pseudomallei S13		Reference isolate	NZ_AAHW00000000
Erwinia carotovora subsp. *atroseptica* SCRI1043		Potato stem	NC_004547
Haemophilus ducreyi 35000HP		Human	NC_002940
Haemophilus influenzae 1056†	ICE*Hin1056*	Human blood	AJ627386
Histophilus somni 129PT		Animal mucous membranes	NZ_AABO00000000
Histophilus somni 2336		Pneumoniac calf	NZ_AACJ00000000
Legionella pneumophila Paris		Endemic isolate	NC_006368
Legionella pneumophila Lens		Epidemic isolate	NC_006369
Legionella pneumophila subsp. *pneumophila* Philadelphia 1		Lung	NC_002942
Neisseria gonorrhoeae MS11	GGI	Human	AY803022
Nitrosomonas eutropha C71		Soil	NZ_AAJE00000000
Photorhabdus luminescens subsp. *laumondii* TT01		Nematode symbiont	NC_005126
Pseudomonas aeruginosa C‡	PAGI-2	Cystic fibrosis patient	AF440523
Pseudomonas aeruginosa C	pKLC102	Cystic fibrosis patient	AY257538
Pseudomonas aeruginosa C	PAGI-5	Cystic fibrosis patient	Unpublished
Pseudomonas aeruginosa PA14	PAPI-1	Clinical isolate	AY273869
Pseudomonas aeruginosa SG17M	PAGI-3	River	AF440524
Pseudomonas aeruginosa TB	PAGI-6	Cystic fibrosis patient	Unpublished
Pseudomonas fluorescens Pf-5		Plant commensal	NC_004129
Pseudomonas fluorescens PfO-1		Agricultural soil	NZ_AAAT00000000
Pseudomonas fluorescens SBW25		Rhizosphere	Unpublished
Pseudomonas sp. B13§	*clc* transposon	Soil	AJ617740
Pseudomonas syringae pv. phaseolicola 1302A	PPHGI-1	Bean plant	AJ870974
Pseudomonas syringae pv. syringae B728a		Snap bean leaflet	NC_007005
Pseudomonas syringae pv. tomato DC3000		Tomato plant	NC_004578
Rubrivivax gelatinosus PM1 (*Methylobium petroleophilum* PM1)		Water pollution treatment plant	NZ_AAEM00000000
Salmonella enterica subsp. *enterica* serovar Typhi CT18	SPI-7	Typhoid fever patient	NC_003198
Xanthomonas axonopodis pv. citri 306		Citrus plant	NC_003919
Xanthomonas campestris pv. campestris 8004		Cauliflower	NC_007086
Xylella fastidiosa 9a5c		Orange plant	NC_002488
Yersinia enterocolitica 8081		Gastroenteritis	Unpublished
Yersinia pseudotuberculosis 32777	YAPI	Infection	AJ627388

*Two related islands (a and b) are present in the genome (see Fig. 2).

†Nearly identical islands are present in *H. influenzae* strains R2866 and 86-028NP.

‡An identical island is present in *Cupriavidus metallidurans* CH34.

§100 % sequence identity with *clc* plus further genes in *Ralstonia* sp. JS705 and *Burkholderia xenovorans* LB400.

region (Fig. 3a, b). The *clc* element was found to be capable to self-transfer to other beta- and gammaproteobacteria (Ravatn *et al.*, 1998), whereby it excised from the donor chromosome at a low frequency and self-transferred to the new host in which it reintegrated. *clc*-like elements have a worldwide distribution; for example, an element was detected in a *Ralstonia* sp. isolate from contaminated groundwater in Texas (Müller *et al.*, 2003). PAGI-2 also seems to be widely distributed in diverse species, geographical regions and habitats. PAGI-2 was originally detected in the lungs of a German cystic fibrosis patient (Larbig *et al.*, 2002) who has been chronically carrying the genome island stably integrated into the chromosome for more than 20 years. Despite its apparently stable chromosomal integration, PAGI-2 was found to be widespread in the *P. aeruginosa* population (see above), indicating that it can be mobilized and transferred to other strains. This hypothesis is supported by the fact that a copy of PAGI-2 was identified with 100 % nucleotide sequence identity in the recently sequenced *Cupriavidus* (formerly *Ralstonia*) *metallidurans* CH34 chromosome. This strain was isolated in 1976 from the sludge of a zinc decantation tank in Belgium that was polluted with high concentrations of several heavy metals. A small collection of *C. metallidurans* and *Cupriavidus campiniensis* environmental isolates was screened for the presence of PAGI-2 by hybridization onto the PAGI-2 macroarray. Three of five *C. metallidurans* strains (Fig. 3c) and both of two *C. campiniensis* strains harboured close homologues to all PAGI-2 ORFs in their chromosomes. One *C. metallidurans* strain was carrying a truncated version of PAGI-2 (J. Klockgether, unpublished data). All *Cupriavidus* strains were isolated from polluted environmental habitats. In summary, closely related *clc*-like and PAGI-2-like genome islands (Fig. 2) were identified in clinical and environmental beta- and gammaproteobacteria of diverse geographical

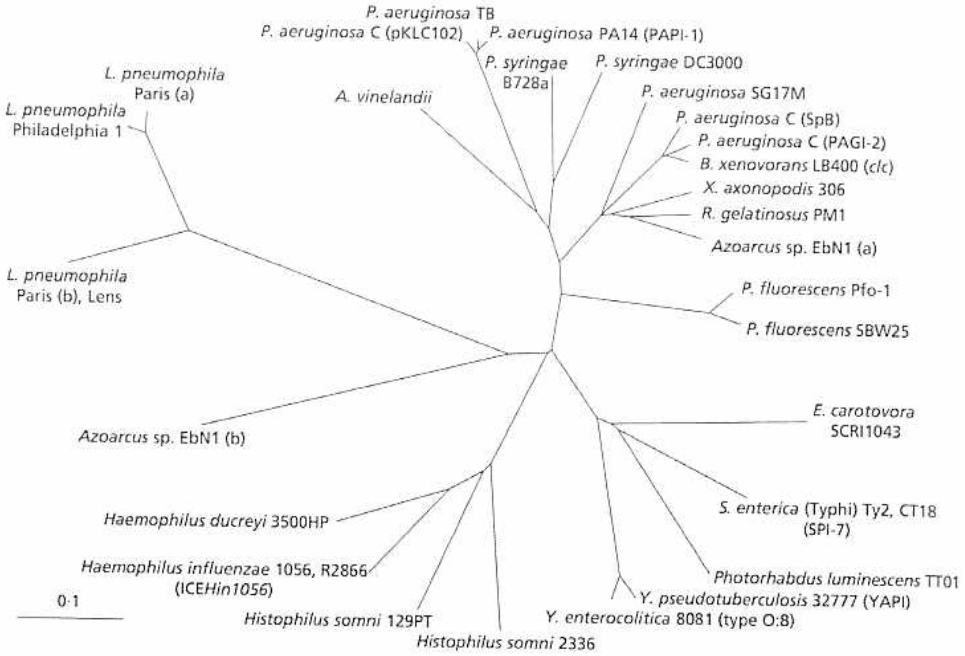


Fig. 2. Unrooted tree based on CLUSTAL X 1.83 analysis of 32 GEIs. The amino acid sequences of all homologues of the *P. aeruginosa* C ORF47 (Larbig *et al.*, 2002) were aligned. Each strain containing a GEI was identified and, where available, the designation of the GEI (e.g. PAPI-1) is given in parentheses. See Table 1 for strain and sequence details.

origin. These GEIs are made up of a module that encodes strain-specific features and a module of conserved syntenic homologues, the majority of which are conserved hypotheticals of yet unknown function. The syntenic gene contig is probably involved in the excision, transfer, integration and stabilization of the genome island (Sentchilo *et al.*, 2003a, b).

CONCLUSIONS

The gene repertoire of a bacterial cell consists of genes that have been transmitted vertically over long periods of time and genes that were acquired or generated at various points of the lineage, including some very recently (Lerat *et al.*, 2005). Horizontal gene transfer provides most of the diversity in the genomic repertoire, but the majority of the horizontally acquired genes that persist in genomes are transmitted strictly vertically (Lerat *et al.*, 2005). Hence, despite substantial horizontal gene transfer, the phylogenetic relationships between taxa are robust, as indicated by the congruence of gene trees based on rRNA gene sequences, gene contents or average amino acid identity of shared genes (Konstantinidis & Tiedje, 2005). This chapter reports on an exception to this rule. GEIs are part of the variable bacterial gene pool. They are only present in

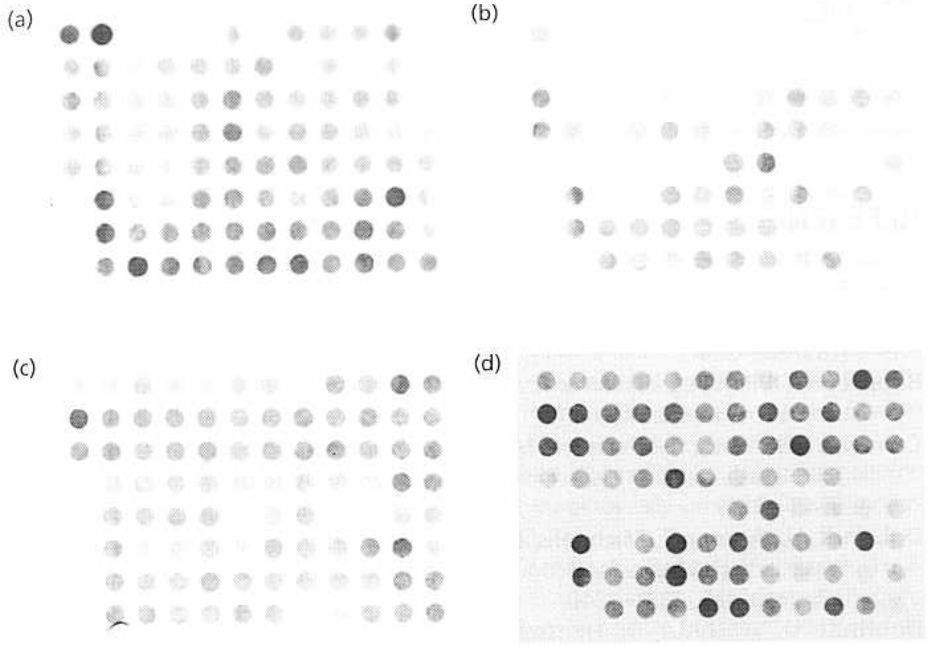


Fig. 3. Hybridization of the PAGI-2 macroarray with genomic DNA from clinical and environmental isolates. (a) *P. aeruginosa* strain C, the initial source of PAGI-2, isolated from the airways of a patient with cystic fibrosis in Hannover, Germany (Larbig *et al.*, 2002). (b) *Ralstonia* sp. strain JS705, isolated from groundwater at Kelly Air Force Base, near San Antonio, TX, USA, which uses chlorobenzene as sole carbon and energy source (Müller *et al.*, 2003). (c) *Cupriavidus metallidurans* KT21, a wastewater isolate from Germany. (d) *C. metallidurans* CH79, recovered from polluted sediment at a zinc factory in Liège, Belgium. Strain JS705 was obtained from Jan Roelof van der Meer (University of Lausanne, Switzerland); the two *C. metallidurans* strains were provided by Max Mergeay (Belgian Nuclear Research Centre, Mol, Belgium).

some strains or clones of a species and hence contribute to intra- and interclonal diversity. Moreover, some island types have been identified in numerous taxa. The most prominent example known to date is an evolutionarily ancient GEI that is widespread among beta- and gammaproteobacteria and has been identified in isolates from clinical and environmental habitats. This GEI consists of two modules: one module endows strain-specific features and the other consists of a set of conserved syntenic genes. Sequence identity of the syntenic set is less than that of vertically transmitted genes, indicating that these genes and probably also the encoded functions are more diverse than those encoded by orthologues of core genomes. Most genes of the conserved module are classified as conserved hypotheticals of yet unknown function, although at least a subset should be involved in the excision, transfer, integration or stabilization of the island (Sentchilo *et al.*, 2003a, b). Future research will unravel whether the syntenic gene set is not only essential for the maintenance of the GEI, but also influences the expression, regulation and function of gene products of the core genome.

ACKNOWLEDGEMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft (Europäisches Graduiertenkolleg 'Pseudomonas: Pathogenicity and Biotechnology' and the Schwerpunktprogramm 'Ökologie bakterieller Krankheitserreger - molekulare und evolutionäre Aspekte').

REFERENCES

Bach, S., de Almeida, A. & Carniel, E. (2000). The *Yersinia* high-pathogenicity island is present in different members of the family *Enterobacteriaceae*. *FEMS Microbiol Lett* **183**, 289–294.

Bopp, L. H. (1986). Degradation of highly chlorinated PCBs by *Pseudomonas* strain LB400. *J Ind Microbiol* **1**, 23–29.

Carniel, E., Guilvout, I. & Prentice, M. (1996). Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. *J Bacteriol* **178**, 6743–6751.

Dobrindt, U., Agerer, F., Michaelis, K. & 7 other authors (2003). Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* **185**, 1831–1840.

Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414–424.

Feil, H., Feil, W. S., Chain, P. & 17 other authors (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. *Proc Natl Acad Sci U S A* **102**, 11064–11069.

Frantz, B. & Chakrabarty, A. M. (1987). Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation. *Proc Natl Acad Sci U S A* **84**, 4460–4464.

He, J., Baldini, R. L., Deziel, E. & 7 other authors (2004). The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc Natl Acad Sci U S A* **101**, 2530–2535.

Holden, M. T., Titball, R. W., Peacock, S. J. & 45 other authors (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **101**, 14240–14245.

Kingsley, R. A., van Amsterdam, K., Kramer, N. & Bäumler, A. J. (2000). The *shdA* gene is restricted to serotypes of *Salmonella enterica* subspecies I and contributes to efficient and prolonged fecal shedding. *Infect Immun* **68**, 2720–2727.

Klockgether, J., Reva, O., Larbig, K. & Tümmler, B. (2004). Sequence analysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. *J Bacteriol* **186**, 518–534.

Konstantinidis, K. T. & Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**, 6258–6264.

Larbig, K. D., Christmann, A., Johann, A., Klockgether, J., Hartsch, T., Merkl, R., Wiehlmann, L., Fritz, H. J. & Tümmler, B. (2002). Gene islands integrated into $tRNA^{Gly}$ genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol* **184**, 6665–6680.

Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**, e130.

Liang, X., Pham, X. Q., Olson, M. V. & Lory, S. (2001). Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. *J Bacteriol* **183**, 843–853.

Mohd-Zain, Z., Turner, S. L., Cerdeno-Tarraga, A. M. & 7 other authors (2004). Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. *J Bacteriol* **186**, 8114–8122.

Morales, G., Wiehlmann, L., Gudowius, P., van Delden, C., Tümmler, B., Martinez, J. L. & Rojo, F. (2004). Structure of *Pseudomonas aeruginosa* populations analyzed by single nucleotide polymorphism and pulsed-field gel electrophoresis genotyping. *J Bacteriol* **186**, 4228–4237.

Müller, T. A., Werlen, C., Spain, J. & van der Meer, J. R. (2003). Evolution of a chlorobenzene degradative pathway among bacteria in a contaminated groundwater mediated by a genomic island in *Ralstonia*. *Environ Microbiol* **5**, 163–173.

Nakazawa, T. (2002). Travels of a *Pseudomonas*, from Japan around the world. *Environ Microbiol* **4**, 782–786.

Nelson, K. E., Weinel, C., Paulsen, I. T. & 40 other authors (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**, 799–808.

Oelschlaeger, T. A., Zhang, D., Schubert, S., Carniel, E., Rabsch, W., Karch, H. & Hacker, J. (2003). The high-pathogenicity island is absent in human pathogens of *Salmonella enterica* subspecies I but present in isolates of subspecies III and VI. *J Bacteriol* **185**, 1107–1111.

Pickard, D., Wain, J., Baker, S. & 12 other authors (2003). Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. *J Bacteriol* **185**, 5055–5065.

Ravatn, R., Zehnder, A. J. & van der Meer, J. R. (1998). Low-frequency horizontal transfer of an element containing the chlorocatechol degradation genes from *Pseudomonas* sp. strain B13 to *Pseudomonas putida* F1 and to indigenous bacteria in laboratory-scale activated-sludge microcosms. *Appl Environ Microbiol* **64**, 2126–2132.

Reva, O. N. & Tümmler, B. (2004). Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns. *BMC Bioinformatics* **5**, 90.

Reva, O. N. & Tümmler, B. (2005). Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. *BMC Bioinformatics* **6**, 251.

Römling, U., Greipel, J. & Tümmler, B. (1995). Gradient of genomic diversity in the *Pseudomonas aeruginosa* chromosome. *Mol Microbiol* **17**, 323–332.

Römling, U., Schmidt, K. D. & Tümmler, B. (1997). Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J Mol Biol* **271**, 386–404.

Schubert, S., Rakin, A., Karch, H., Carniel, E. & 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.

Schubert, S., Dufke, S., Sorsa, J. & Heesemann, J. (2004). A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the *Yersinia* high-pathogenicity island. *Mol Microbiol* **51**, 837–848.

Sentchilo, V., Ravatn, R., Werlen, C., Zehnder, A. J. & van der Meer, J. R. (2003a). Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. strain B13. *J Bacteriol* **185**, 4530–4538.

Sentchilo, V., Zehnder, A. J. & van der Meer, J. R. (2003b). Characterization of two

Pseudomonas sp. strain B13. *Mol Microbiol* **49**, 93–104.

Springael, D., Peys, K., Ryngaert, A. & 8 other authors (2002). Community shifts in a seeded 3-chlorobenzoate degrading membrane biofilm reactor: indications for involvement of in situ horizontal transfer of the *clc*-element from inoculum to contaminant bacteria. *Environ Microbiol* **4**, 70–80.

van der Meer, J. R. & Sentchilo, V. (2003). Genomic islands and the evolution of catabolic pathways in bacteria. *Curr Opin Biotechnol* **14**, 248–254.

Welch, R. A., Burland, V., Plunkett, G., III & 16 other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**, 17020–17024.

Evolving gene clusters in soil bacteria

Alice Morningstar, William H. Gaze, Sahar Tolba and Elizabeth M. H. Wellington

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

INTRODUCTION

Soil is heterogeneous in nearly all respects and contains a huge diversity of micro-organisms. The availability of carbon and other energy sources, mineral nutrients and water varies considerably over space and time, as does temperature. Adaptations to nutrient poverty including oligotrophy and zymogeny (upsurge in growth when nutrients are available) are common. The water films essential for microbial life in soil are discontinuous, and only clay particles have the necessary charges to hold water against the pull of gravity. Clay-coated soil particles cluster together to form aggregates, and these aggregates or clusters of aggregates with their adjacent water form the microhabitats in which bacteria function (Stotzky, 1997). The result of the discrete microhabitats in soil is that microbial population dynamics and interactions are very different from those in well-mixed substrates such as some aquatic environments. Soil is also a reservoir for pesticides and other chemical and microbiological inputs from slurry application, all of which will have a selective impact on the indigenous bacteria.

Bacterial evolutionary histories are difficult to untangle. Different scales of evolution occur simultaneously, from events possible over a few generations (chromosomal rearrangement, gene deletion and acquisition of genes via horizontal transfer) to the eon-scale generative evolution which creates diversity from which the novel functional genes of the future will be selected. In the age of genomics we are developing the tools to study the ecology of microbes in soil. The few metagenomic projects undertaken so far illustrate the diversity of bacteria in soil (>3000 ribotypes in a Minnesota farm soil

SGM symposium 66: Prokaryotic diversity: mechanisms and significance.
Editors N. A. Logan, H. M. Lappin-Scott & P. C. F. Oyston. Cambridge University Press. ISBN 0 521 86935 8 ©SGM 2006