Whole genome plasticity in pathogenic bacteria

Ulrich Dobrindt* and Jörg Hacker†

The exploitation of bacterial genome sequences has so far provided a wealth of new general information about the genetic diversity of bacteria, such as that of many pathogens. Comparative genomics uncovered many genome variations in closely related bacteria and revealed basic principles involved in bacterial diversification, improving our knowledge of the evolution of bacterial pathogens. A correlation between metabolic versatility and genome size has become evident. The degenerated life styles of obligate intracellular pathogens correlate with significantly reduced genome sizes, a phenomenon that has been termed 'evolution by reduction'. These mechanisms can permanently alter bacterial genotypes and result in adaptation to their environment by genome optimization. In this review, we summarize the recent results of genome-wide approaches to studying the genetic diversity of pathogenic bacteria that indicate that the acquisition of DNA and the loss of genetic information are two important mechanisms that contribute to strain-specific differences in genome content.

Addresses
Institut für Molekulare Infektionsbiologie der Universität Würzburg, Röntgenring 11, 97070 Würzburg, Germany
*e-mail: ulrich.dobrindt@mail.uni-wuerzburg.de
†e-mail: j.hacker@mail.uni-wuerzburg.de
Corresponding author: Jörg Hacker

Current Opinion in Microbiology 2001, 4:550–557

1369-5274/01/$ – see front matter
© 2001 Elsevier Science Ltd. All rights reserved.

Abbreviations

EHEC	enterohaemorrhagic *E. coli*
GEI	genomic island
HPI	high pathogenicity island
ORF	open reading frame
PAI	pathogenicity island

Introduction

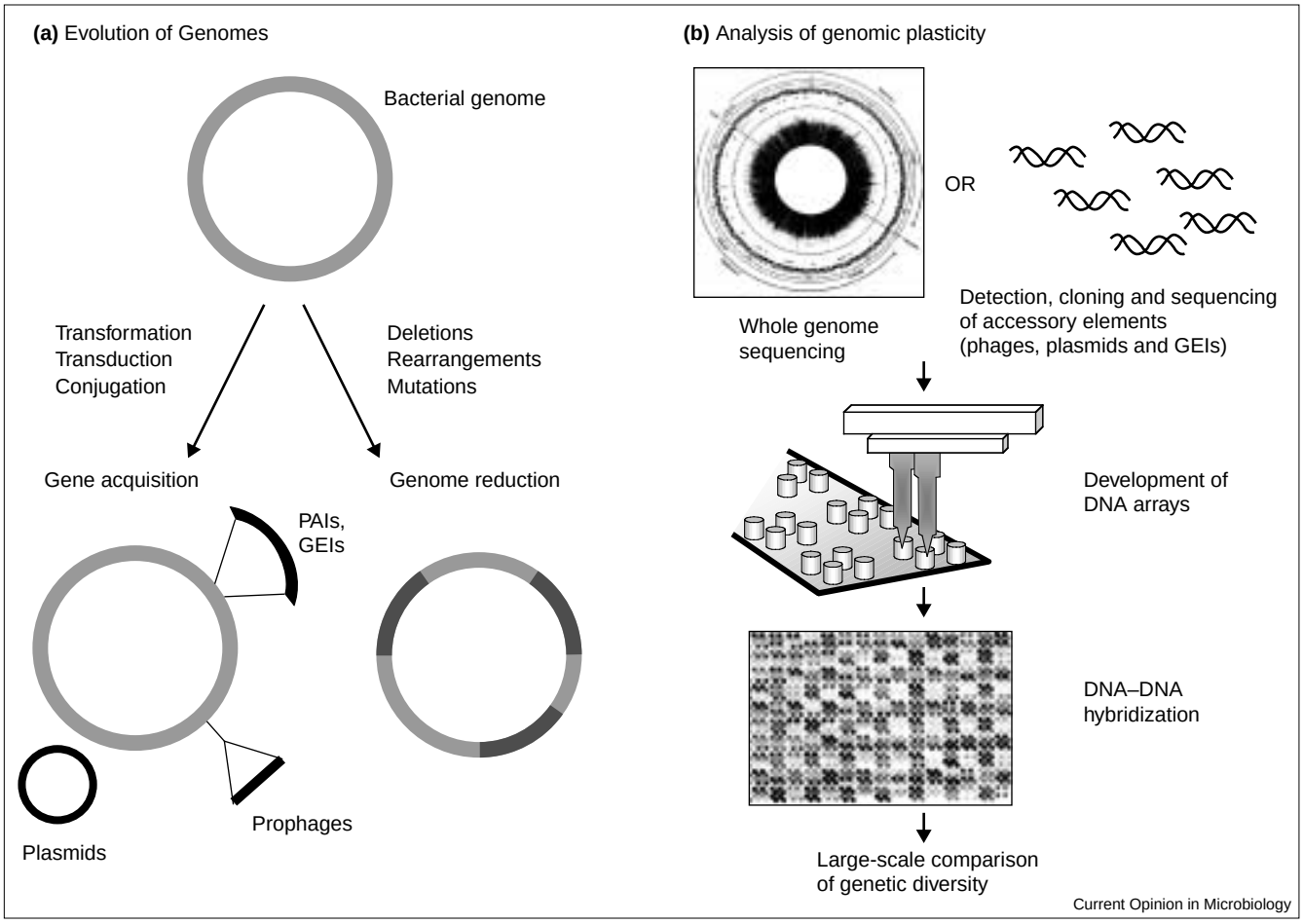
It is well known that the broad spectrum of physiological and virulence properties of bacterial pathogens mirrors the existence of different subsets of genes enabling the pathogen's different lifestyles. During the past few years, it has become evident that one strain's genome sequence is not entirely representative for other members of the species. The use of 'whole genome approaches' to study the genetic diversity of closely related bacterial strains increases rapidly. Whereas the comparison of complete genome sequences is time consuming and not favourable to compare genomes of several strains of one species, DNA–DNA hybridization using DNA arrays provides a cheaper and faster alternative to assessing the genome content of closely related strains. The results obtained so far from comparative genomics underline, first, the importance of gene acquisition and loss for genome evolution and, second, that the genetic organization reflects the bacterial lifestyle. In this review, we summarize recent data of comparative genomics and discuss the general mechanisms involved in genome plasticity and in the evolution of bacterial species or variants.

Genome organization of pathogenic bacteria – gene acquisition and gene loss

The sizes and organization of bacterial genomes vary considerably. The smallest and largest genomes of bacterial pathogens known so far are those of *Mycoplasma genitalium* (580 kb) and *Pseudomonas aeruginosa* (6300 kb), respectively. There is significant variation in genome size within bacterial genera and species. Different numbers and different combinations of circular or linear chromosomes and extrachromosomal linear or circular replicons exist in bacterial pathogens [1–4]. Bacterial genomes can generally be considered to be composed of a universally present 'core' of genes providing the backbone of genetic information that should be conserved in most of the bacteria. This also includes the 'minimal set', as has been previously described [5,6]. The minimal gene set is shared by the vast majority of bacteria and forms the basic composition of the core gene pool. However, it is hard to characterize because it depends on the growth conditions specific to the individual bacterial species and because there is no clear definition of the minimal requirements and processes of life. The genome also contains a flexible gene pool consisting of an 'assortment' of strain-specific genetic information that may provide additional properties enabling these species to adapt to special environmental conditions. A correlation between metabolic versatility and genome size has become evident. The large genome size of *P. aeruginosa* (larger than that of, for example, *Escherichia coli*) results from great genetic complexity and mirrors its greater capacity to adapt to its environment [7]. Differences in genome size among bacterial species reflect size variations of the flexible gene pool and are mainly due to the acquisition and loss of genomic DNA. DNA uptake may involve several mechanisms and results in gain of additional genetic information (Figure 1a).

Conversely, because of close adaptation of parasites and intracellular pathogens (e.g. *Mycoplasma* spp., *Chlamydia trachomatis*, *Mycobacterium leprae* and *Rickettsia prowazekii*) to the physiologically stable environments of their host cells, a reductive genome evolution occurred that led to the loss of genes not essential for life within the host (a phenomenon termed 'evolution by reduction'), but leaving genes responsible for functions needed for the performance of key metabolic steps and for interconversions of metabolites obtained from the host. A typical feature of 'evolution by reduction' is a considerably high fraction of pseudogenes within genomes [8–11]. Other bacterial species or pathotypes (i.e. strains of a bacterial species that cause the

Figure 1



Bacterial genome plasticity: evolutionary mechanisms and large-scale analysis. (a) Genome evolution is mainly due to the acquisition and loss of genetic information. Several mechanisms can result in DNA uptake or in the reduction of genome content (shaded areas). Mobile genetic elements such as plasmids, bacteriophages and pathogenicity islands play an important role in DNA acquisition. (b) DNA array-based analysis of genome plasticity requires sequence information obtained by complete genome sequencing or by subcloning of DNA regions of interest. On the basis of sequence information, DNA arrays can be developed that facilitate large-scale comparison of genetic diversity.

same type of infection, for example, uropathogenic *E. coli* (UPEC) that cause urinary tract infection, enterohaemorrhagic *E. coli* (EHEC) that cause haemorrhagic colitis and haemolytic uraemic syndrome, etc.) that underwent genome content reduction during their evolution include *Yersinia pseudotuberculosis*, from which *Yersinia pestis* emerged, and non-pathogenic *E. coli*, from which *Shigella* spp. emerged. Several genes present in *Y. pseudotuberculosis* are absent or non-functional in *Y. pestis* [12•]. When pathogenic *Shigella* spp. arose from a non-pathogenic *E. coli* ancestor, the loss of the *ompT* and *cadA* ('black hole') genes contributed to their evolution [13,14]. The deletion of a large chromosomal region ('black hole') present in about 90 % of all *E. coli* strains but absent in enteroinvasive *E. coli* and *Shigella* strains promotes the virulence of these intestinal pathogens. This chromosomal region contains several genes, including *cadA*, which codes for the enzyme lysine decarboxylase and *ompT*. The lack of *cadA* and *ompT* supports the virulence of *Shigella* and enteroinvasive *E. coli*, as cadaverine, a product of the CadA-catalyzed enzymatic reaction, inhibits the enterotoxin activity of these strains, and as the expression of the *ompT*-encoded protease interferes with the VirG protein required for intercellular spread.

Genetic mechanisms leading to genome plasticity

The tremendous genetic diversity observed among pathogenic bacteria mirrors their different lifestyles and physiological versatilities and evolves from adaptation to their niches or growth conditions. Although the chromosomal organization of more closely related bacteria is more conserved, considerable genome variability exists within different genera and among different isolates of a single bacterial species. For example, the chromosomes of different enterobacterial species share a highly homologous common backbone that is interrupted by many variable regions [15–19]. Several mechanisms are involved in this adaptive evolution (Figure 1a, Table 1). Point mutations

Table 1

Mechanisms contributing to genome plasticity.

Genetic element or mechanism	Consequences
Gain of properties	
Point mutation	Alteration of gene expression
Homologous recombination	DNA rearrangements, inversion, duplication, deletion of DNA Integration of horizontally acquired DNA
Transformation	Gain of additional genetic information
IS elements, composite transposons	Insertion, deletion, inversion of DNA, alteration of gene expression
Integrons	Transfer of genes, DNA rearrangements
Conjugative transposons, plasmids	Conjugation Horizontal gene transfer Mobilization of other plasmids
Bacteriophages	Generalized or specialized transduction Horizontal gene transfer
GEIs or PAIs, pathogenicity islets	Horizontal gene transfer Integration and deletion of large DNA regions
Loss of properties	
Point mutation	Alteration of gene expression, loss of function
Homologous recombination	DNA rearrangements, deletion of DNA, integration of horizontally acquired DNA
Transposition	Alteration of gene expression, loss of function

GEI, genomic island; IS, insertion; PAI, pathogenicity island.

contribute to genetic variation. Homologous recombination creates DNA rearrangements, e.g. deletions, duplications and inversions. Chromosomal rearrangements have been frequently observed during the genome comparison of different isolates of several bacterial species [7,10,16,17,20]. Numerous repetitive elements within the genomes of *Neisseria meningitidis* strains are regarded to be involved in antigenic variation and genome fluidity, as they may facilitate recombination [21,22]. Homologous recombination also affects genome organization by mediating the integration of horizontally acquired mobile genetic elements [23]. Transformation of DNA generally increases the gene content.

The amount of laterally acquired genetic information is unexpectedly high in many bacterial pathogens [24,25••]. The great importance of horizontal gene transfer for the evolution of genetic diversity is underlined by the substantial number of functional and non-functional accessory genetic elements (plasmids, bacteriophages and genomic islands [GEIs]) detected in genomes of different pathogenic bacteria. These accessory elements are important 'carriers' of genetic information containing many genes coding for pathogenic features of many bacteria. Transposable genetic elements (e.g. insertion [IS] elements, transposons and integrons) contribute to genome variability by transposition events [26] and can be involved in mobilization of genetic information. Obligate intracellular parasites tend to contain fewer accessory and repetitive DNA elements than do extracellular pathogens. This is probably because of their isolation from genetic exchange with other bacteria [9,27]. In the genomes of the different *C. trachomatis* and *C. pneumoniae* strains sequenced so far, no IS elements or other dispersed repeated sequences that may be involved in chromosomal rearrangements have been detected. In the case of *M. leprae*, only three bacteriophage-related genes and 26 transposase gene fragments have been identified within the genome, whereas the genome of *M. tuberculosis* strain H37Rv contains at least two prophages and 56 intact or truncated IS elements. In addition, multiple copies of dispersed repetitive sequences (RLEP, REPLEP and LEPREP) are present within the genome of *M. leprae* TN. They exhibit some features of classical IS elements and are several hundred basepairs in size. These elements are considered to be involved in genomic reorganization.

GEIs represent an important type of flexible genetic element. It is assumed that mobile genetic elements were involved in the evolution of many GEIs. Some of them are still mobile elements. GEIs represent 'foreign' DNA entities that can be subdivided according to their size and encoded functions. They are at least 10 kb in size. DNA regions smaller than 10 kb are considered to be genomic islets. These elements can provide a certain advantage for the host cell under specific conditions. If virulence-associated genes are located on such a genomic element (Table 2), it is considered to be a pathogenicity island (PAI) [28]. Some PAIs represent laterally transferable DNA regions. The so-called 'high pathogenicity island' (HPI) initially discovered in *Yersinia* spp. encodes a siderophore system. The HPI core element has been detected in many other enterobacteria [29,30] and represents a 'broad-host-range GEI' that contributes to adaptation to different growth conditions. This DNA element can 'move' within the chromosome between the three different asparaginyl-tRNA genes of *Y. pestis* and *Y. pseudotuberculosis*. The HPI consists of a highly conserved

Table 2

Examples of mobile genetic elements of pathogenic bacteria.

Species or genetic element	Designation	Virulence factors or functions
Pathogenicity islands		
UPEC	PAI I_{536}–PAI III_{536}, PAI I_{J96}–PAI II_{J96}, PAI I_{CFT073}	Adhesins, α-hemolysin, cytotoxic necrotizing factor 1, heat-resistant agglutinin 1, siderophore system
EPEC	LEE	Adhesin, type III secretion
	espC-PAI	EspC protein (enterotoxin)
Vibrio cholerae O1, O139	VPI	Pili, regulation
Staphylococcus aureus	TSST-1-PAI (SaPI1, SaPI2, SaPIbov)	TSST-1
	Exotoxin-PAI	Exotoxin
	Enterotoxin-PAI	Enterotoxin, serine protease
Pathogenicity islets		
E. coli, *Shigella dysenteriae*	*chuA* or *shuA* locus	Heme uptake
Salmonella enterica sv. Typhimurium	*msgA*/*pagC* locus	Outer-membrane protein, intramacrophage survival
Streptococcus pyogenes	*vir* region	M-like proteins, proteases
Plasmids		
Extraintestinal *Escherichia coli*	pHly152, Vir plasmid, pCol	α-Hemolysin, cytotoxic necrotizing factor 2, aerobactin (iron uptake)
Intestinal *E. coli*	pO157, EAF plasmid, Vir plasmids	Adhesins, LT, ST enterotoxins, catalase, hemolysin
Shigella flexneri	pWR100, pWR501	Invasins, type III secretion proteins, enterotoxin
Clostridium tetani	pCL1	Tetanus toxin
Bacteriophages		
Clostridium botulinum	cI	Botulinum toxin A, B
Corynebacteria diphtheriae	β	Diphtheria toxin A, B
E. coli (EHEC)	H19, 933	Shiga toxin A, B
S. aureus	ϕ42	Enterotoxin A, B
V. cholerae	CTXϕ	Cholera toxin A, B

EAF, EPEC adherence factor; EHEC, enterohaemorrhagic *E. coli*; EPEC, enteropathogenic *E. coli*; LEE, locus of enterocyte effacement; LT, heat-labile enterotoxin; PAI, pathogenicity island; SaPI, *S. aureus* pathogenicity island; ST, heat-stable enterotoxin; TSST-1; toxic-shock toxin 1; UPEC, uropathogenic *E. coli*; VPI, *V. cholerae* pathogenicity island.

core and a host-strain-dependent, variable part [31–33]. The locus of enterocyte effacement (LEE)-PAI of EHEC and enteropathogenic *E. coli* (EPEC) strains has been independently acquired by different evolutionary lineages of *E. coli* and exhibits structural variations, depending on the lineage [34,35••,36]. The genetic structure of the *Shigella* island 2 (SHI2)-PAI is also variable in *Shigella* spp. [37,38]. Thus, GEIs and PAIs contribute to inter- and intraspecific variability in genome content [23].

Plasmids contribute to genome plasticity because of their transferability and their ability to mobilize other co-resident plasmids. In addition, these usually extrachromosomal elements can be chromosomally integrated [39–42]. Many virulence-associated determinants and antibiotic resistance genes are located on plasmids (Table 2). Genetic exchange between plasmids and other replicons is a constantly ongoing process [24,43], exemplified by the similarities between PAI-specific regions of *E. coli* and *Shigella flexneri* virulence plasmids.

Bacteriophages play a major role in the evolution of pathogenic bacteria. Important virulence genes are bacteriophage-encoded (Table 2). PAI-like structures in *Vibrio cholerae* and *Staphylococcus aureus* (designated *V. cholerae* pathogenicity island [VPI] and *S. aureus* pathogenicity island 1 [SaPI1], respectively) can be mobilized by propagation of the helper phage CP-T1 (in the case of the VPI), or of the phages 13 and 80α (in the case of the SaPII), suggesting that they represent defective prophages [44,45]. Helper-phage-mediated mobilization of DNA regions could be of general importance for genome plasticity, as functional bacteriophage integrase genes or corresponding pseudogenes are located on many GEIs. Expression of an integrase gene may mobilize several chromosomal regions within one bacterial strain. For example, the defective prophage P4 can be mobilized by the P2 helper phage [46]. These broad-host-range phages can be detected in other enterobacteria. Several enterobacterial PAIs contain a functional or non-functional integrase gene of the P4 family [47]. Thus, induction of a P2 phage or expression of a P4 integrase could result in the excision of multiple chromosomal regions.

Comparative genomics of pathogenic bacteria

Many bacterial populations are clonal. *E. coli* isolates can be grouped into particular clones that have started to evolve under competition as distinct genetic types. They arose in parallel by loss of and by ordered gain of genetic information and are maintained during adaptation to their niches.

Table 3

Comparison of genetic diversity among closely related bacteria.

Reference species or strain	Investigated species or strain	Percentage of ORFs missing from the reference strain	References
Escherichia coli MG1655 (K12)	*E. coli* EDL933 (EHEC)	12	[17]
E. coli MG1655 (K12)	*E. coli* 536 (UPEC)	9	[53]
E. coli MG1655 (K12)	*E. coli* IHE3034 (MENEC)	5	[53]
E. coli MG1655 (K12)	*Klebsiella pneumoniae* 342	27	[54]
Streptococcus pneumoniae KNR.7/87	*S. pneumoniae* strains (n=20)	10*	[55]
Helicobacter pylori strains J99 and 26695	*H. pylori* strains (n=15)	22*	[59•]
Mycobacterium tuberculosis H37Rv	*M. tuberculosis* strains (n=19)	0.3†	[60]
M. tuberculosis H37Rv	*Mycobacterium bovis*	2.3*	[50]
M. tuberculosis H37Rv	*M. bovis* BCG strains (n=13)	3.3*	[50]

*Percentage of ORFs that are missing in at least one of the investigated strains. †Average number of *M. tuberculosis* H37Rv-specific ORFs missing in the tested strains. EHEC, enterohaemorrhagic *E. coli*; MENEC, meningitis-causing *E. coli*; UPEC, uropathogenic *E. coli*.

Because of horizontal gene transfer, their further evolution is constantly in progress [25••,35]. However, in certain bacterial species (*Streptococcus pneumoniae*, *Streptococcus pyogenes*, *N. meningitidis* and *S. aureus*), an unusually high frequency of recombination results in disintegration of the 'clonal frame' [48]. In *P. aeruginosa* populations, a high recombination frequency leads to random distribution of alleles and formation of hierarchically equivalent genotypes. These genotypes show no adaptation to their habitats and exhibit such versatility that their colonization of different niches does not require specialization [49].

DNA microarrays facilitate the investigation of the genetic diversity among different bacterial strains (Figure 1b, Table 3). This technology was used for the first time to compare several *Mycobacterium bovis* vaccine strains. Several chromosomal deletions in different vaccine strains were discovered in comparison to their progenitor, these deletions possibly being responsible for the variable effectiveness of the BCG vaccine [50]. DNA microarray technology and techniques for comparing complete genome sequences have now been used several times to analyze the genome content and genetic variability of different bacterial species. Interestingly, the analysis of complete genome sequences of a non-pathogenic *E. coli* and two EHEC strains revealed that a surprisingly great proportion of the flexible gene pool consists of uncharacterized open reading frames (ORFs) without any obvious function (about 38% of the *E. coli* K12 strain MG1655-specific ORFS, and 35% of the EHEC strain EDL933-specific ORFs).

Other major constituents of the flexible gene pool are mobile and accessory genetic elements or non-functional fragments of these elements [15–17,51]. Comparison of the genetic diversity of different non-pathogenic *E. coli* strains using *E. coli* MG1655-specific DNA macroarrays confirmed that chromosomal variations exist all over the chromosome and that the flexible gene pool consists of a high number of putative ORFs and mobile genetic elements [52•]. Using DNA macroarrays, DNA–DNA hybridization of genomic DNA of seven different pathogenic *E. coli* isolates with that of the non-pathogenic *E. coli* K12 strain MG1655 demonstrated that 5–10% of all translatable *E. coli* strain MG1655-specific ORFs were not detectable in the different pathogenic *E. coli* isolates. About 50–60% of these ORFs can be functionally grouped as hypothetical and unclassified. ORFs categorized as components of mobile genetic elements also showed a high variability with respect to their presence in these strains (U Dobrindt, unpublished data; [53]). An *E. coli* K12-specific DNA microarray was also used to analyze the gene content of the closely related maize endophyte, *Klebsiella pneumoniae* 342. Approximately 70% of the *E. coli* K12-specific ORFs were found to be at least 55% identical to ORFs of this *Klebsiella* strain [54] (see also Update). A *S. pneumoniae*-specific DNA microarray was used to compare the genome contents of 20 *S. pneumoniae* isolates. The presence of up to 10% of the genes varied in different strains tested. Seven out of ten large gene clusters belonging to this group of variable genomic regions are associated with transposases and one resembles a prophage, indicating that they represent putative mobile genetic elements [55]. The comparison of complete genome sequences of a pathogenic and a non-pathogenic *S. aureus* strain showed that a considerable fraction of genes seemed to be horizontally acquired. Many virulence-associated as well as antibiotic resistance genes were located on mobile accessory DNA elements [24]. These data stress the importance of horizontal gene transfer and mobile genetic elements for genetic diversity among pathogenic bacteria.

Physical genome organization and genome content in terms of functionality in the naturally competent organism *H. pylori* is largely conserved. According to two complete *H. pylori* genome sequences, approximately 6% of the genes are strain-specific. Only two chromosomal regions, the PAI and the so-called plasticity zone (a chromosomal region that exhibits a higher rate of DNA reorganization than the rest of the chromosome), were identified to be hot spots for strain-specific genes that exhibit several features,

indicating that lateral gene transfer or extensive recombinational events had occurred. However, in spite of this functional and organizational analogy, identical gene sequences in *H. pylori* strains are extremely rare. This extensive nucleotide diversity underlines the importance of extremely frequent recombination for diversity in this species [55-58]. A DNA array was developed on the basis of the completely sequenced genomes of two *H. pylori* strains and was used for genome analysis of 15 *H. pylori* clinical isolates. In contrast to the results obtained from the comparison of the two available complete *H. pylori* genome sequences, the DNA array-based genome comparison revealed that about 22% of *H. pylori* genes were not detectable in all strains tested. The majority of them represents genes that encode unknown proteins or proteins that affect genetic exchange [59•].

Similarly, two *C. trachomatis* and two *C. pneumoniae* genomes were found to be almost identical in gene content and gene order. Marked differences with respect to gene order and genome organization between *C. trachomatis* and *C. pneumoniae* as well as between the genomes of the two strains of these *Chlamydia* species tested were only visible near the origins of replication and termination where chromosomal inversions occurred. This chromosomal region, termed 'plasticity zone', exists in both species and shows a higher degree of recombinational activity than the rest of the chromosome. The mechanisms responsible for rearrangements in the plasticity zone are unclear [27].

Conclusions

Comparative genomics contributes significantly to our understanding of bacterial evolution and bacterial pathogenesis. The increasing knowledge of genetic diversity of bacteria obtained from comparative genome analyses indicates that bacterial genomes are constantly evolving structures. Mechanisms contributing to genome plasticity include point mutations, DNA rearrangements and horizontal gene transfer. The acquisition by a variety of bacteria of foreign DNA coding for novel phenotypes is an important factor in the evolution of bacterial pathogens. Mobile genetic elements are frequently involved in horizontal gene transfer of virulence-associated genes. Acquisition of genetic information as a result of lateral spread of mobile genetic elements contributes to constant and rapid evolution of bacterial pathogens. Additionally, genetic diversity arises from deletion and loss of chromosomal regions. The larger genome complexity of physiologically versatile bacteria, in comparison to the reductive genome evolution of parasitic or symbiotic bacteria, suggests that the genome content mirrors the lifestyles of these bacteria, and results from the ongoing process of genome optimization.

Update

Recent work by Akman and Akso [61] has demonstrated that DNA arrays may also be used for assessing data on the genome content and general biology of closely related bacteria in the absence of entire genome sequences. In order to gain information on genome composition and general gene expression patterns of *Wigglesworthia glossinidia* (a non-culturable obligate endosymbiont of the tsetse fly and a member of the Enterobacteriaceae), an *E. coli* K-12-specific macroarray was used for genome comparison and expression analysis. According to DNA–DNA hybridization experiments, the *Wigglesworthia* genome size was estimated to be smaller than 770 kb and about 650 orthologues of *E. coli* genes were detected (approximately 85% of the entire genome). The expression profiles suggest that this endosymbiont may be a facultative anaerobic bacterium that utilizes ammonia as a major nitrogen source.

Acknowledgements

Our own work related to this topic was supported by the Deutsche Forschungsgemeinschaft (SFB 479) and the 'Fonds der Chemischen Industrie'.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bergthorsson U, Ochman H: Distribution of chromosome length variation in natural isolates of ***Escherichia coli***. *Mol Biol Evol* 1998, 15:6-16.

2. Casjens S: The diverse and dynamic structure of bacterial genomes. *Annu Rev Genet* 1998, 32:339-377.

3. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L *et al.*: DNA sequence of both chromosomes of the cholera pathogen ***Vibrio cholerae***. *Nature* 2000, 406:477-483.

4. Michaux S, Paillisson J, Carles-Nurit MJ, Bourg G, Allardet-Servent A, Ramuz M: Presence of two independent chromosomes in the ***Brucella melitensis*** 16M genome. *J Bacteriol* 1993, 175:701-705.

5. Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC: Global transposon mutagenesis and a minimal ***Mycoplasma*** genome. *Science* 1999, 286:2165-2169.

6. Mushegian AR, Koonin EV: A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 1996, 93:10268-10273.

7. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M *et al.*: Complete genome sequence of ***Pseudomonas aeruginosa*** PA01, an opportunistic pathogen. *Nature* 2000, 406:959-964.

8. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG: The genome sequence of ***Rickettsia prowazekii*** and the origin of mitochondria. *Nature* 1998, 396:133-140.

9. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D *et al.*: Massive gene decay in the leprosy bacillus. *Nature* 2001, 409:1007-1011.

10. Himmelreich R, Plagens H, Hilbert H, Reiner B, Herrmann R: Comparative analysis of the genomes of the bacteria ***Mycoplasma pneumoniae*** and ***Mycoplasma genitalium***. *Nucleic Acids Res* 1997, 25:701-712.

11. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q *et al.*: Genome sequence of an obligate intracellular pathogen of humans: ***Chlamydia trachomatis***. *Science* 1998, 282:754-759.

12. • Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E: ***Yersinia pestis***, the cause of plague, is a recently emerged clone of ***Yersinia pseudotuberculosis***. *Proc Natl Acad Sci USA* 1999, 96:14043-14048.

In this paper, the evolutionary relationship between *Y. pseudotuberculosis* and *Y. pestis* is described on the basis of the genetic differences and genetic variability between the three pathogenic species of *Yersinia*. The impact of the genetic differences observed between *Y. pseudotuberculosis* and *Y pestis* on their virulence is discussed.

13. Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A: 'Black holes' and bacterial pathogenicity: a large genomic deletion that enhances the virulence of ***Shigella*** spp. and enteroinvasive ***Escherichia coli***. *Proc Natl Acad Sci USA* 1998, 95:3943-3948.

14. Nakata N, Tobe T, Fukuda I, Suzuki T, Komatsu K, Yoshikawa M, Sasakawa C: The absence of a surface protease, OmpT, determines the intercellular spreading ability of ***Shigella***: the relationship between the ***ompT*** and ***kcpA*** loci. *Mol Microbiol* 1993, 9:459-468.

15. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Villes J, Glasner JD, Rode CK, Mayhew GF: The complete genome sequence ***of Escherichia coli*** K-12. *Science* 1997, 277:1453-1462.

16. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T *et al.*: Complete genome sequence of enterohemorrhagic ***Escherichia coli*** O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 2001, 8:11-22.

17. Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA *et al.*: Genome sequence of enterohaemorrhagic ***Escherichia coli*** O157:H7. *Nature* 2001, 409:529-533.

18. Pupo GM, Lan R, Reeves PR: Multiple independent origins of ***Shigella*** clones of ***Escherichia coli*** and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 2000, 97:10567-10572.

19. Wong RM, Wong KK, Benson NR, McClelland M: Sample sequencing of a ***Salmonella typhimurium*** LT2 lambda library: comparison to the ***Escherichia coli*** K12 genome. *FEMS Microbiol Lett* 1999, 173:411-423.

20. Jackson RW, Mansfield JW, Arnold DL, Sesma A, Paynter CD, Murillo J, Taylor JD, Vivian A: Excision from tRNA genes of a large chromosomal region, carrying ***avrPphB***, associated with race change in the bean pathogen, ***Pseudomonas syringae*** pv. ***phaseolicola***. *Mol Microbiol* 2000, 38:186-197.

21. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T *et al.*: Complete DNA sequence of a serogroup A strain of ***Neisseria meningitidis*** Z2491. *Nature* 2000, 404:502-506.

22. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ *et al.*: Complete genome sequence of ***Neisseria meningitidis*** serogroup B strain MC58. *Science* 2000, 287:1809-1815.

23. Hacker J, Kaper JB: Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 2000, 54:641-679.

24. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki KI, Nagai Y *et al.*: Whole genome sequencing of methicillin-resistant ***Staphylococcus aureus***. *Lancet* 2001, 357:1225-1240.

25. •• Ochman H, Lawrence JG, Groisman EA: Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000, 405:299-304.

This is a detailed review of the importance of horizontal gene transfer for the adaptive evolution of different pathogenic bacteria.

26. Craig NL: Transposition. In: *Escherichia coli and Salmonella typhimurium*. Edited by Neidhardt FC, Curtis III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE. Washington DC: ASM Press; 1996:2339-2362.

27. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K *et al.*: Genome sequences of ***Chlamydia trachomatis*** MoPn and ***Chlamydia pneumoniae*** AR39. *Nucl Acids Res* 2000, 28:1397-1406.

28. Kaper JB, Hacker J: *Pathogenicity Islands, Plasmids and Other Mobile Elements*. Washington, DC: ASM Press; 1999.

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

30. Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, Ölschläger T, Hacker J: A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing ***Escherichia coli*** clonal lineages. *Infect Immun* 1999, 67:5994-6001.

31. Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E: The high-pathogenicity island of ***Yersinia pseudotuberculosis*** can be inserted into any of the three chromosomal ***asn*** tRNA genes. *Mol Microbiol* 1998, 30:965-978.

32. Hare JM, Wagner AK, McDonough KA: Independent acquisition and insertion into different chromosomal locations of the same pathogenicity island in ***Yersinia pestis*** and ***Yersinia pseudotuberculosis***. *Mol Microbiol* 1999, 31:291-303.

33. Rakin A, Noelting C, Schubert S, Heesemann J: Common and specific characteristics of the high-pathogenicity island of ***Yersinia enterocolitica***. *Infect Immun* 1999, 67:5265-5274.

34. Perna NT, Mayhew GF, Posfai G, Elliott S, Donnenberg MS, Kaper JB, Blattner FR: Molecular evolution of a pathogenicity island from enterohemorrhagic ***Escherichia coli*** O157:H7. *Infect Immun* 1998, 66:3810-3817.

35. •• Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS: Parallel evolution of virulence in pathogenic ***Escherichia coli***. *Nature* 2000, 406:64-67.

This paper provides an important demonstration that horizontal gene transfer and the ordered acquisition of genetic information can lead to parallel evolution of pathogenic bacteria.

36. Zhu C, Agin TS, Elliott SJ, Johnson LA, Thate TE, Kaper JB, Boedeker EC: Complete nucleotide sequence and analysis of the locus of enterocyte Effacement from rabbit diarrheagenic ***Escherichia coli*** RDEC-1. *Infect Immun* 2001, 69:2107-2115.

37. Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA: The ***selC***-associated SHI-2 pathogenicity island of ***Shigella flexneri***. *Mol Microbiol* 1999, 33:74-83.

38. Vokes SA, Reeves SA, Torres AG, Payne SM: The aerobactin iron transport system genes in ***Shigella flexneri*** are present within a pathogenicity island. *Mol Microbiol* 1999, 33:63-73.

39. Kiewitz C, Larbig K, Klockgether J, Weinel C, Tümmler B: Monitoring genome evolution ***ex vivo***: reversible chromosomal integration of a 106 kb plasmid at two tRNA(Lys) gene loci in sequential ***Pseudomonas aeruginosa*** airway isolates. *Microbiology* 2000, 146:2365-2373.

40. Lüneberg E, Mayer B, Daryab N, Kooistra O, Zähringer U, Rohde M, Swanson J, Frosch M: Chromosomal insertion and excision of a 30 kb unstable genetic element is responsible for phase variation of lipopolysaccharide and other virulence determinants in ***Legionella pneumophila***. *Mol Microbiol* 2001, 39:1259-1271.

41. Protsenko OA, Filippov AA, Kutyrev VV: Integration of the plasmid encoding the synthesis of capsular antigen and murine toxin into ***Yersinia pestis*** chromosome. *Microb Pathog* 1991, 11:123-128.

42. Zagaglia C, Casalino M, Colonna B, Conti C, Calconi A, Nicoletti M: Virulence plasmids of enteroinvasive ***Escherichia coli*** and ***Shigella flexneri*** integrate into a specific site on the host chromosome: integration greatly reduces expression of plasmid-carried virulence genes. *Infect Immun* 1991, 59:792-799.

43. Brunder W, Karch H: Genome plasticity in ***Enterobacteriaceae***. *Int J Med Microbiol* 2000, 290:153-165.

44. Karaolis DK, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR: A ***Vibrio cholerae*** pathogenicity island associated with epidemic and pandemic strains. *Proc Natl Acad Sci USA* 1998, 95:3134-3139.

45. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP: The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in ***Staphylococcus aureus***. *Mol Microbiol* 1998, 29:527-543.

46. Christie GE, Calendar R: Interactions between satellite bacteriophage P4 and its helpers. *Annu Rev Genet* 1990, 24:465-490.

47. Hacker J, Blum-Oehler G, Janke B, Nagy G, Goebel W: Pathogenicity islands of extraintestinal ***Escherichia coli***. In *Pathogenicity Islands and Other Mobile Virulence Elements*. Edited by Kaper JB, Hacker J. Washington, DC: ASM Press; 1999:59-76.

48. Feil EJ, Holmes EC, Bessen DE, Chan MS, Day NP, Enright MC, Goldstein R, Hood DW, Kalia A, Moore CE *et al.*: Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* 2001, 98:182-187.

49. Kiewitz C, Tümmler B: Sequence diversity of ***Pseudomonas aeruginosa***: impact on population structure and genome evolution. *J Bacteriol* 2000, 182:3125-3135.

50. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM: Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999, 284:1520-1523.

51. Lawrence JG, Ochman H: Molecular archaeology of the ***Escherichia coli*** genome. *Proc Natl Acad Sci USA* 1998, 95:9413-9417.

52. • Ochman H, Jones IB: Evolutionary dynamics of full genome content in ***Escherichia coli***. *EMBO J* 2000, 19:6637-6643.
This paper reports the use of a DNA array for the analysis of genetic diversity and evolutionary relationship of five non-pathogenic *E. coli* strains. It describes for the first time the use of DNA arrays for the characterization of the genetic diversity and evolutionary history among selected non-pathogenic *E. coli* strains.

53. Dobrindt U, Hentschel U, Kaper JB, Hacker J: Genome plasticity in pathogenic and non-pathogenic enterobacteria. *Curr Top Microbiol Immunol* 2001, 264:in press.

54. Dong Y, Glasner JD, Blattner FR, Triplett EW: Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, ***Klebsiella pneumoniae*** 342, by microarray hybridization with ***Escherichia coli*** K-12 open reading frames. *Appl Environ Microbiol* 2001, 67:1911-1921.

55. Hakenbeck R, Balmelle N, Weber B, Gardes C, Keck W, de Saizieu A: Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of ***Streptococcus pneumoniae***. *Infect Immun* 2001, 69:2477-2486.

56. Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A, van Doorn LJ: Recombination and clonal groupings within ***Helicobacter pylori*** from different geographical regions. *Mol Microbiol* 1999, 32:459-470.

57. Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL *et al.*: Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen ***Helicobacter pylori***. *Nature* 1999, 397:176-180.

58. Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M: Free recombination within ***Helicobacter pylori***. *Proc Natl Acad Sci USA* 1998, 95:12619-12624.

59. • Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S: A whole-genome microarray reveals genetic diversity among ***Helicobacter pylori*** strains. *Proc Natl Acad Sci USA* 2000, 97:14668-14673.
This paper reports that genome comparison of 15 *H. pylori* isolates using a DNA array revealed a greater genetic diversity than was assumed on the basis of two complete *H. pylori* genome sequences.

60. Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, Small PM: Comparing genomes within the species ***Mycobacterium tuberculosis***. *Genome Res* 2001, 11:547-554.

61. Akman L, Aksoy S: A novel application of gene arrays: ***Escherichia coli*** array provides insight into the biology of the obligate endosymbiont of tsetse flies. *Proc Natl Acad Sci USA* 2001, 98:7546-7551.