

Part I

Basic Principles

1

How Bacterial Pathogens were Constructed

Ulrich Dobrindt and Jörg Hacker

1.1

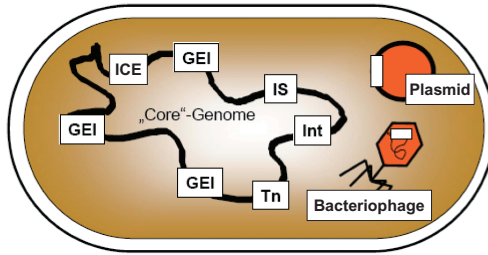
Introduction

The complete genome sequences of multiple variants of almost all bacterial pathogens have already been published. There is large variation in size and content of bacterial genomes between different genera and species but also among strains of the same species. The adaptive capability ('versatility') of bacteria directly correlates with genome size [1]. Genome optimization in bacteria with an intracellular parasitic lifestyle, for example, *Chlamydia trachomatis* and *Rickettsia prowazekii* (genome size 1.04 Mb and 1.11 Mb, respectively) which live in a constant and rich environment, involves marked genome reduction due to the loss of genetic information coding for traits and functions that could be obtained from the host.

Consequently, the genomes of such bacteria living in close association with their host are much smaller than those of environmental bacteria that may be facultative pathogens, e.g. *Pseudomonas aeruginosa* (genome size 6.3 Mb), as these organisms have to live under different and variable growth conditions.

The analysis of these sequences revealed that genomes of pathogenic microbes, but also of microbes from other sources, may comprise different numbers of circular or linear chromosomes, extrachromosomal linear or circular replicons as well as different combinations thereof. The genome of *Vibrio cholerae*, the causative agent of cholera, contains two chromosomes, both of them encode important functions [2, 3]. Most prokaryotes, however, contain a single circular chromosome which comprises genes providing the backbone of genetic information required for essential cellular processes and which is not transferable *per se*. The arrangement of genes on the chromosome is characterized by a frequently clustered organization or a close link between functionally related genes. Genes located on the 'core' part of the chromosome exhibit a relatively homogenous G+C content and a specific codon usage. The genomic organization of closely related bacteria is very similar [4].

The structure of the 'core' regions of the genome, however, is often interrupted by the presence of DNA stretches harboring genes with a G+C content and a codon



Plasmids
 Bacteriophages
 Genomic islands (GEI)
 Integrative and conjugative elements (ICE)
 Integrations (Int)
 (Conjugative)Transposons (Tn), IS elements (IS)

Figure 1.1 Composition of a bacterial genome. The core genome is shown together with the various elements of the flexible gene pool.

usage which differs from those of the 'core' genome. This 'flexible' gene pool consists of strain-specific 'assortments' of genetic information mainly represented by mobile genetic elements, such as plasmids, bacteriophages, genomic/pathogenicity islands (GEIs/PAIs), integrons, IS-elements and transposons (see Figure 1.1 and Table 1.1), which provide additional traits contributing to the adaptation of microbes under certain environmental conditions, e.g. resistance to antibiotics, production of toxic compounds as well as other virulence factors [5].

1.2 Composition of the Flexible Gene Pool

IS elements and transposons are considered to be 'jumping genes' as they can frequently change their chromosomal localization. Generally, transposable elements fall into three groups: (i) insertion sequences (relatively small mobile genetic entities that generally encode no functions other than those involved in their mobility and exhibit short terminal inverted repeat sequences), (ii) composite transposons (flanked by insertion sequences at both ends), and (iii) non-composite transposons (lack flanking insertion sequences). Many transposons carry antibiotic resistance genes [6].

Integrations can also jump and may carry determinants for antibiotic resistances and other properties. They can be considered as a natural genetic engineering system because they are assembly platforms, i.e. DNA elements that acquire genes embedded in exogenous gene cassettes and convert them to functional genes by ensuring their correct expression. They were initially identified because of their important role in the spread of antibiotic-resistance determinants. Meanwhile,

Table 1.1 Mobilizable and transferable accessory genetic elements of the flexible gene pool that contribute to genome plasticity.

DNA element	Size	Characteristics
Pathogenicity/ Genomic Island	20–200kb	Modular, mosaic-like composition, site-specific insertion into tRNA genes
Plasmids	2–200kb	Modular, mosaic-like composition, sometimes chromosomal insertion possible, can transfer transposons and IS elements
Integrative and conjugative elements	30–150kb	Site-specific insertion into tRNA genes, modular composition
Prophages	40–100kb	Site-specific insertion frequently into tRNA genes
(Conjugate) Transposons	20–100kb	Mainly in Gram-positive bacteria
(Super-)Integrans	~100kb	Mainly in Gammaproteobacteria, many promoter-less gene cassettes
Transposons	2.5–25kb	Flanked by IS elements
Insertion sequences	0.7–2.5kb	Encode only their integrase

larger integron structures, termed superintegrons, have been discovered which further support their importance in bacterial genome evolution. The latter contain hundreds of accessory genes and constitute a significant fraction of the genomes of many bacterial species including important intestinal pathogens such as *Vibrio cholerae* [7].

Transferable genetic elements such as bacteriophages and plasmids can function as vehicles laterally transporting genetic information, thus playing an important role in bacterial evolution. Bacteriophages may carry genes that bring about new functions or modify existing ones upon becoming part of the host genome. Bacteriophages often carry toxin genes, e.g. the cholera toxin determinant, the genes encoding diphtheria toxin or the Shiga toxins. Plasmids are circular, self-replicating DNA molecules. They are autonomous molecules and exist in cells as extrachromosomal genomes, although some plasmids can be inserted into a bacterial chromosome. The role of transferable elements as vectors as well as the constantly ongoing recombination between different mobilizable and transferable DNA elements is exemplified for example by the description of the association of multi-drug resistance and virulence determinants in form of a large virulence plasmid in multi-drug-resistant *Salmonella enterica* serotype Typhimurium that carries several virulence genes and two class 1 integrons [8] or a multi-resistance plasmid of *Klebsiella pneumoniae* that resembles enterobacterial integrative and conjugative elements as well as plasmids found in *Yersinia pestis* [9, 10].

Genomic islands (GEIs) are large chromosomal sectors which are part of the flexible gene pool as they represent formerly transferred DNA regions. Pathogenicity islands (PAIs) are particular GEIs present in pathogenic bacteria but absent in the majority of closely related non-pathogenic variants. They carry one or more virulence-associated gene(s) and are frequently associated with tRNA genes as well as flanked by repeat structures. As PAIs are often unstable and as they contain mobility genes coding for integrases or transposases they contribute to the dynamic character of chromosomes in pathogens [5, 11]. Certain conjugative and self-transmissible genetic elements exhibit features of plasmids and bacteriophages as they can be transferred by conjugation and can integrate into the bacterial chromosome. Such integrative and conjugative elements (ICEs) may function as progenitors of some PAIs/GEIs during genome evolution [12].

Accessory genetic elements like transposons, integrons, insertion elements and genomic or pathogenicity islands (GEIs, PAIs) represent major constituents of the flexible gene pool. The genomic islands are seldom fixed but rather bear the potential for ongoing rearrangements, deletions and insertions. As a result, the stable chromosomal backbone and the flexible gene pool are constantly undergoing repeated insertions and deletions leading to new variants, pathotypes and over a long-term process to new species. Thus, such genomes are composed of clonally evolving DNA regions that are periodically disrupted due to exchange of already existing gene blocks by homologous recombination, and insertion of horizontally acquired DNA segments. The majority of strain- or pathotype-specific regions have accumulated over time by repeated horizontal gene transfer, frequently with successive transfers of different elements into identical loci of the core chromosome. The existence of so many different horizontally-acquired sequences in genomic islands differentiating closely related strains indicates that many of them are only temporarily present in the genome or provide a specific advantage to the individual lifestyle of particular strains.

1.3 Mechanisms Involved in Genome Dynamics

Different mechanisms contribute synergistically to dynamic bacterial genome evolution: first, point mutations can be accumulated and result in the diversification of genes. Second, variation in the bacterial genome organization results from transposition, site-specific as well as homologous recombination of DNA regions. Repeated sequences, notably mobile DNA elements, play a major role in the overall genome plasticity as they facilitate recombination and these DNA rearrangements may cause relocation or deletion of genomic regions. DNA recombination also causes gene duplication. Duplicated genes can evolve as orthologues or by divergent evolution as paralogues. Additionally, foreign genetic material is acquired by horizontal gene transfer. This acquisition of genetic material results in extremely dynamic genomes in which substantial amounts of DNA are introduced into and deleted from the chromosome. The extent of gene acquisition in bacteria differs

considerably between different genera and species and may be responsible for up to 15% of the complete genome sequence [13]. The combination of such evolutionary mechanisms during prokaryotic genome evolution can be nicely illustrated by the shuffling and disruption of operons in many bacteria due to rearrangements and gene transfer [14].

1.4

Bacterial Genome Optimization Using *Escherichia coli* as a Model

The species *Escherichia coli* comprises pathogenic and non-pathogenic variants [5, 15, 16]. Pathogenic isolates may cause infections of the intestine as well as extraintestinal infections such as urinary tract infections, sepsis and meningitis. Commensal *E. coli* have the capacity to colonize the intestine of humans and many animals without causing any harm.

The pheno- and genotypic variability of pathogenic and commensal *E. coli* correlate with their genome content. *E. coli* genomes vary in size from 4.6 to 5.6 Mb [17]. These size differences among individual *E. coli* genomes indicate the presence of different amounts of strain-specific genetic information, which may represent up to 30% of the complete genome content. Genes for many virulence traits as well as antibiotic resistance genes of IPEC (intestinal porcine epithelial cell line-1) and ExPEC, especially those characteristic of the different pathotypes, may be encoded on mobile and accessory genetic elements, e.g. genomic and pathogenicity islands or transposons [5, 18–21], plasmids and bacteriophages [22–26].

Extraintestinal pathogenic *E. coli* (ExPEC) are epidemiologically and phylogenetically distinct from many commensal strains as well as from IPEC. A variety of virulence factors directly contribute to pathotype-specific disease and their distribution is thus restricted to the corresponding pathotypes. For instance, the ETT-1 type III secretion system and its translocated effectors are usually indicative of enterohemorrhagic *E. coli* (EHEC) and enteropathogenic *E. coli* (EPEC). The heat-stable or heat-labile enterotoxins are characteristic of enterotoxigenic *E. coli* (ETEC) [16]. Certain invasion genes such as *ibeA* as well as the K1 capsule determinant are frequently present in invasive ExPEC [27]. In many cases, however, ExPEC and commensal *E. coli* share a large fraction of their genome [15, 28, 29]. There are also many so-called virulence-associated factors in ExPEC such as colicins, certain fimbriae, siderophore systems and toxins [15, 30–32] that have probably evolved to enhance survival in the gut and/or transmission between hosts, and therefore will be shared with some commensal strains and sometimes even with IPEC.

As many *E. coli* virulence-associated genes may be located on transmissible genetic elements such as bacteriophages, plasmids or transposons, individual DNA regions can be exchanged between the chromosome and mobile genetic elements with the capacity to integrate into and excise from the bacterial chromosome. Accordingly, several identical or closely related virulence determinants can be found on the chromosome or on mobile DNA elements.

So-called colicin plasmids represent an interesting example of such mobile elements which in large part exhibit considerable sequence similarity to PAIs in *E. coli* and contribute to PAI evolution and the spread of virulence traits among individual strains. Large colicin plasmids are found primarily in virulent, mainly septicemic *E. coli* strains and they seem to be a characteristic marker for avian pathogenic *E. coli* (APEC), causing systemic infections in poultry for example. The genetic structure of such plasmids from APEC highly resembles that of other large colicin and related plasmids and several PAIs of *E. coli* [33] as they carry several genes that have been previously associated with APEC virulence such as those coding for colicins, toxins and factors involved in serum resistance. It also contains several operons associated with iron acquisition including the aerobactin system (*iuc/iut*), the *iro* determinant coding for salmochelin, the *sit* operon, coding for an ABC transport system involved in iron and manganese transport and the *eitA-D* genes that code for a putative iron transport system. The operons coding for the siderophore systems *sit*, *iut* and *iro* as well as the *iss* gene involved in serum resistance can be found on the bacterial chromosome as well as on the colicin plasmids [34]. In APEC, these determinants are exclusively found on colicin plasmids whereas in other pathogenic enterobacteria they are frequently located on chromosomal PAIs [33, 35].

As already mentioned, many PAI regions exhibit notable homology to fragments of mobile genetic elements such as bacteriophages and virulence plasmids. In addition, multiple copies of accessory DNA elements in one genome facilitate homologous recombination within one or between different islands or horizontally-acquired DNA elements thus leading to rearrangements, deletions and acquisition of 'foreign' DNA. Consequently, many PAIs have a mosaic-like modular structure. Although many of them superficially resemble each other with respect to the presence and/or genetic linkage of certain virulence determinants, PAI composition, structural organization and chromosomal localization can be highly variable even among strains of the same patho- or serotype [36, 37]. A recent comparative genomic and phylogenetic analysis of 20 commensal and pathogenic *E. coli* strains indicated that genome variability within the *E. coli* species is restricted to a small number of conserved chromosomal positions. These hot spots of gene acquisition and loss correspond to regions of abundant and parallel insertions and deletions of DNA and can comprise between 10 and 157 genes [38]. Such large genomic regions seem to be more permissive to the insertion/deletion of accessory DNA elements as once they are chromosomally inserted they serve as preferred regions for other insertion/deletion events. The resulting larger regions are then distributed within *E. coli* by homologous recombination between their flanking genes rather than by dissemination of the individual accessory elements comprising these genomic hot spots [38].

It has been shown that the virulence features of pathogenic microbes may not only vary between different strains, but also between different time points of infection within a single strain. To analyze the flexibility of PAIs, we used the method of 'island probing' [39–41]. Using this method it was possible to calculate the deletion rate of respective PAIs. With a deletion rate of 10^{-3} to 10^{-4} , they can be excised

from their respective genome. The integrases, which are encoded by the respective PAIs are involved in this deletion processes. Generally they act highly specifically at their respective target sites [42, 43], but may sometimes also have the capacity to engage in 'cross-talk' and thus also to be involved in the excision of another island. PAI deletion may be beneficial in the course of an infection as later stages of infection may develop into chronic infections. Under these conditions, several virulence-associated traits, especially toxins, may be disadvantageous.

1.5 Genome Plasticity during Infection

Variation in the bacterial genome organization results in large part from transposition and homologous recombination. Repeated sequences, notably mobile DNA elements, play a major role in genome plasticity. Repeated DNA sequences play a primordial role in the overall genome plasticity as they are frequently involved in recombination events. It is postulated, and also frequently observed, that 'useless' or somehow deleterious genes are lost from bacterial genomes (genome reduction). Without a positive selection pressure, a gene will be lost rapidly. This loss will be even more rapid if the particular gene is part of a mobile DNA element [44].

Recent studies suggest that genome plasticity seems to be accelerated under *in vivo* conditions as it may facilitate adaptation to various host conditions. During infection, a strong selection pressure is exerted on the pathogen, leading to variations in the clonal lineages. Pheno- and genotypic changes have been reported for consecutive *Escherichia coli* isolates from recurrent bacteremia cases. In some patients, *E. coli* isolates from consecutive recurrent bacteremia episodes exhibited an altered ability to express long chain LPS, capsule or flagella [45]. *E. coli*, the major cause of urinary tract infections (UTI), may also cause asymptomatic bacteriuria (ABU), i.e. a carrier state without causing symptoms. This resembles a state of commensalism with a bacterial monoculture rather than a complex flora. Accordingly, ABU is an interesting model in which to study mechanisms of commensalism and the driving forces within the pathogen and host. Geno- and phenotypic analyses of ABU isolates demonstrated that many ABU strains arose from virulent variants by gene loss. It has been suggested that attenuation may constitute a general mechanism for mucosal pathogens to evolve towards commensalism [46].

For example, phage mobilization contributes significantly to genome alteration in *Staphylococcus aureus* cystic fibrosis isolates during infection. Such a genome alteration could be linked to bacteriophage mobilization, phage conversion or deletion. [47]. Similarly, a recent complete sequence and comparative analysis of the genomes of two representative *P. aeruginosa* strains isolated from cystic fibrosis (CF) patients with that of other *P. aeruginosa* isolates indicated that niche adaptation is a major evolutionary force influencing the composition of bacterial genomes. Unlike the genome reduction seen in host-adapted bacterial pathogens, the genetic

capacity of *P. aeruginosa* is determined by the ability of individual strains to acquire or discard genomic segments, giving rise to strains with customized genomic repertoires [48].

The gastric pathogen *Helicobacter pylori* shows immense genetic variability in gene content and at the sequence level within human populations [49–51]. Comparison of the genome content of 21 closely related pairs of isolates taken from the same patient at different time points showed that the great majority of genetic changes were due to homologous recombination. These results suggest that adaptation of *H. pylori* to the host individual may principally occur through sequence changes rather than loss or gain of genes [52]. Nevertheless, comparative genomic analysis of chronic atrophic gastritis (ChAG) *H. pylori* isolates also suggests that certain genes may have been lost or gained during progression to adenocarcinoma. Furthermore, adaptation to ChAG also includes alteration in the expression of genes encoding components of metal uptake and utilization pathways, outer membrane proteins and virulence factors [53].

1.6 Conclusions

Comparative genomics indicate that a permanent process of construction and deconstruction of microbial genomes represents an important mechanism of genome evolution, both in short-term intervals (microevolution) as well as in long-term periods (macroevolution). These processes, which have been analyzed in detail for *E. coli*, are also valid for other pathogenic as well as non-pathogenic microbes. Pathogenicity or genomic islands, may exhibit a particular ‘life cycle’ of deconstruction and reconstruction: accessory genetic elements can integrate into the core genome preferentially in a site-specific manner (see Figure 1.2). Following ‘reduction’, particular islands lose gene clusters whose products do not contribute to better survival and transmission of the strain. Following re-integration of additional transposons or IS-elements, classical PAIs will be generated. Islands can delete from the chromosome, but they have the capacity to re-integrate and transfer. It can therefore be concluded that the analysis of the mechanisms involved in destruction and restructuring of bacterial genomes results in a model of the evolutionary processes of microbes in general.

Acknowledgments

We thank Claudia Borde for help in the preparation of the manuscript. Our own work is supported by the ‘Deutsche Forschungsgemeinschaft’ (SFB 479) and the ‘Bavarian Research Foundation’. This work was carried out within the framework of the European Virtual Institute for Functional Genomics of Bacterial Pathogens (CEE LSHB-CT-2005-512061) and the ERA-NET Pathogenomics project ‘Deciphering the intersection of commensal and extraintestinal pathogenic *E. coli*’ (Grant no. 0313937A).

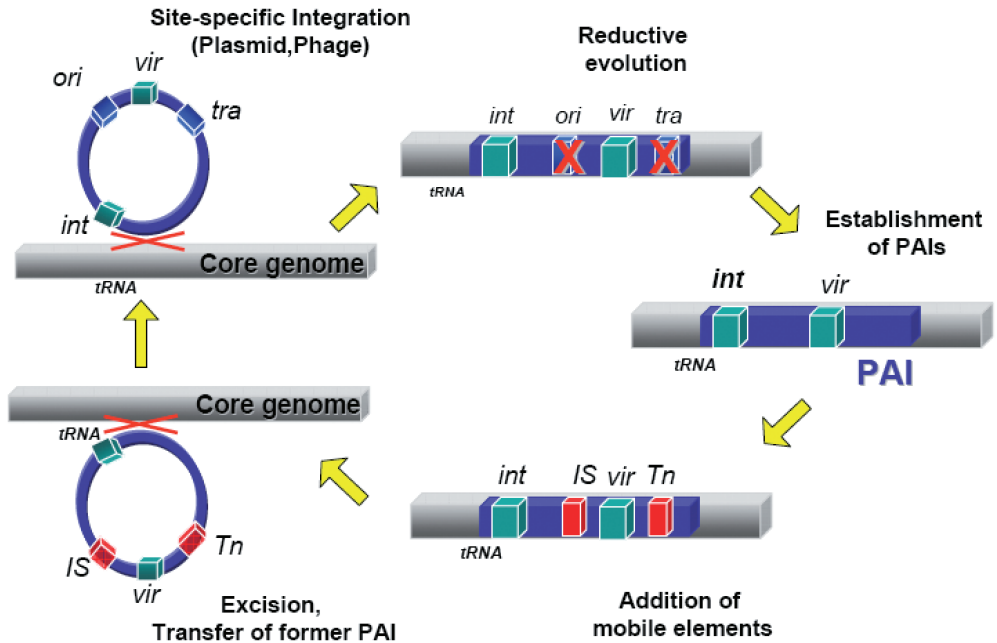


Figure 1.2 Processes of deconstruction and reconstruction of pathogenicity islands. The 'life cycle' shows the development of pathogenicity islands as well as the excision

and integration processes involved. *ori*, origin of replication; *vir*, virulence gene; *tra*, transfer gene; *int*, integrase gene; *IS*, insertion sequence; *Tn*, transposon.

References

- Mira, A., Klasson, L., and Andersson, S.G. (2002) Microbial genome evolution: sources of variability. *Curr. Opin. Microbiol.*, 5, 506–512.
- Casjens, S. (1998) The diverse and dynamic structure of bacterial genomes. *Annu. Rev. Genet.*, 32, 339–377.
- Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umamay, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C., and Fraser, C.M. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*, 406, 477–483.
- Holm, L. (1986) Codon usage and gene expression. *Nucleic Acids Res.*, 14, 3075–3087.
- Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.*, 2, 414–424.
- Craig, N., Craigie, R., Gellert, M., and Lambowitz, A. (2002) *Mobile DNA II*, ASM Press, Washington D.C.
- Mazel, D. (2006) Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.*, 4, 608–620.
- Villa, L., and Carattoli, A. (2005) Integrons and transposons on the *Salmonella enterica* serovar typhimurium

- virulence plasmid. *Antimicrob. Agents Chemother.*, **49**, 1194–1197.
- 9 Lin, T.L., Lee, C.Z., Hsieh, P.F., Tsai, S.F., and Wang, J.T. (2008) Characterization of integrative and conjugative element ICEKp1-associated genomic heterogeneity in a *Klebsiella pneumoniae* strain isolated from a primary liver abscess. *J. Bacteriol.*, **190**, 515–526.
 - 10 Soler Bistué, A.J., Birshan, D., Tomaras, A.P., Dandekar, M., Tran, T., Newmark, J., Bui, D., Gupta, N., Hernandez, K., Sarno, R., Zorreguieta, A., Actis, L.A., and Tolmasky, M.E. (2008) *Klebsiella pneumoniae* multiresistance plasmid pMET1: similarity with the *Yersinia pestis* plasmid pCRY and integrative conjugative elements. *PLoS ONE*, **3**, e1800.
 - 11 Schmidt, H., and Hensel, M. (2004) Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.*, **17**, 14–56.
 - 12 Schubert, S., Dufke, S., Sorsa, J., and 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.
 - 13 Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299–304.
 - 14 Omelchenko, M.V., Makarova, K.S., Wolf, Y.I., Rogozin, I.B., and Koonin, E.V. (2003) Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol.*, **4**, R55.
 - 15 Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., and Dobrindt, U. (2004) Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J. Bacteriol.*, **186**, 5432–5441.
 - 16 Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004) Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.*, **2**, 123–140.
 - 17 Bergthorsson, U., and Ochman, H. (1998) Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol. Biol. Evol.*, **15**, 6–16.
 - 18 Brzuszkiewicz, E., Brüggemann, H., Liesegang, H., Emmerth, M., Ölschläger, T., Nagy, G., Albermann, K., Wagner, C., Buchrieser, C., Emödy, L., Gottschalk, G., Hacker, J., and Dobrindt, U. (2006) How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 12879–12884.
 - 19 Dobrindt, U. (2005) (Patho-)Genomics of *Escherichia coli*. *Int. J. Med. Microbiol.*, **295**, 357–371.
 - 20 Gal-Mor, O., and Finlay, B.B. (2006) Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell. Microbiol.*, **8**, 1707–1719.
 - 21 Nougayrède, J.P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science*, **313**, 848–851.
 - 22 Ogura, Y., Ooka, T., Asadulghani, Terajima, J., Nougayrède, J.P., Kurokawa, K., Tashiro, K., Tobe, T., Nakayama, K., Kuhara, S., Oswald, E., Watanabe, H., and Hayashi, T. (2007) Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. *Genome Biol.*, **8**, R138.
 - 23 Ohnishi, M., Terajima, J., Kurokawa, K., Nakayama, K., Murata, T., Tamura, K., Ogura, Y., Watanabe, H., and Hayashi, T. (2002) Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc. Natl. Acad. Sci. USA*, **99**, 17043–17048.
 - 24 Perna, N.T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*, **409**, 529–533.

- 25 Tobe, T., Beatson, S.A., Taniguchi, H., Abe, H., Bailey, C.M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., and Pallen, M.J. (2006) An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl. Acad. Sci. USA*, **103**, 14941–14946.
- 26 Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A.K., Taboada, E., and Gannon, V.P. (2007) Genome evolution in major *Escherichia coli* O157:H7 lineages. *BMC Genomics*, **8**, 121.
- 27 Moulin-Schouleur, M., Reperant, M., Laurent, S., Bree, A., Mignon-Grasteau, S., Germon, P., Rasschaert, D., and Schouler, C. (2007) Extraintestinal pathogenic *Escherichia coli* strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. *J. Clin. Microbiol.*, **45**, 3366–3376.
- 28 Hejnova, J., Dobrindt, U., Nemcova, R., Rusniok, C., Bomba, A., Frangeul, L., Hacker, J., Glaser, P., Sebo, P., and Buchrieser, C. (2005) Characterization of the flexible genome complement of the commensal *Escherichia coli* strain A0 34/86 (O83: K24: H31). *Microbiology*, **151**, 385–398.
- 29 Rasko, D.A., Rosovitz, M.J., Myers, G.S., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., Henderson, I.R., Sperandio, V., and Ravel, J. (2008) The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.*, **190**, 6881–6893.
- 30 Grozdanov, L., Zähringer, U., Blum-Oehler, G., Brade, L., Henne, A., Knirel, Y.A., Schombel, U., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Rietschel, E.T., and Dobrindt, U. (2002) A single nucleotide exchange in the *wzy* gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of *Escherichia coli* strain Nissle 1917. *J. Bacteriol.*, **184**, 5912–5925.
- 31 Janka, A., Bielaszewska, M., Dobrindt, U., Greune, L., Schmidt, M.A., and Karch, H. (2003) Cytolethal distending toxin gene cluster in enterohemorrhagic *Escherichia coli* O157:H- and O157:H7: characterization and evolutionary considerations. *Infect. Immun.*, **71**, 3634–3638.
- 32 Rendon, M.A., Saldana, Z., Erdem, A.L., Monteiro-Neto, V., Vazquez, A., Kaper, J.B., Puente, J.L., and Giron, J.A. (2007) Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 10637–10642.
- 33 Johnson, T.J., Johnson, S.J., and Nolan, L.K. (2006) Complete DNA sequence of a ColBM plasmid from avian pathogenic *Escherichia coli* suggests that it evolved from closely related ColV virulence plasmids. *J. Bacteriol.*, **188**, 5975–5983.
- 34 Johnson, T.J., Siek, K.E., Johnson, S.J., and Nolan, L.K. (2006) DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J. Bacteriol.*, **188**, 745–758.
- 35 Johnson, T.J., Wannemuehler, Y.M., and Nolan, L.K. (2008) Evolution of the *iss* gene in *Escherichia coli*. *Appl. Environ. Microbiol.*, **74**, 2360–2369.
- 36 Dobrindt, U., Blum-Oehler, G., Nagy, G., Schneider, G., Johann, A., Gottschalk, G., and Hacker, J. (2002) Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. *Infect. Immun.*, **70**, 6365–6372.
- 37 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.
- 38 Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles,

- A.M., Johnson, J., Le Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Medigue, C., Rocha, E.P., and Denamur, E. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.*, **5**, e1000344.
- 39 Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., Gottschalk, G., Carniel, E., and Hacker, J. (2006) Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Mol. Microbiol.*, **61**, 584–595.
- 40 Middendorf, B., Hochhut, B., Leipold, K., Dobrindt, U., Blum-Oehler, G., and Hacker, J. (2004) Instability of pathogenicity islands in uropathogenic *Escherichia coli* 536. *J. Bacteriol.*, **186**, 3086–3096.
- 41 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.
- 42 Turner, S.A., Luck, S.N., Sakellaris, H., Rajakumar, K., and Adler, B. (2004) Role of *attP* in integrase-mediated integration of the *Shigella* resistance locus pathogenicity island of *Shigella flexneri*. *Antimicrob. Agents Chemother.*, **48**, 1028–1031.
- 43 Wilde, C., Mazel, D., Hochhut, B., Middendorf, B., Le Roux, F., Carniel, E., Dobrindt, U., and Hacker, J. (2008) Delineation of the recombination sites necessary for integration of pathogenicity islands II and III into the *Escherichia coli* 536 chromosome. *Mol. Microbiol.*, **68**, 139–151.
- 44 Ahmed, N., Dobrindt, U., Hacker, J., and Hasnain, S.E. (2008) Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat. Rev. Microbiol.*, **6**, 387–394.
- 45 Olesen, B., Kolmos, H.J., Ørskov, F., and Ørskov, I. (1998) *Escherichia coli* bacteraemia in patients with and without haematological malignancies: a study of strain characters and recurrent episodes. *J. Infect.*, **36**, 93–100.
- 46 Zdziarski, J., Svanborg, C., Wullt, B., Hacker, J., and Dobrindt, U. (2008) Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infect. Immun.*, **76**, 695–703.
- 47 Goerke, C., Matias, S., Papenberg, Y., Dasbach, S., Dietz, K., Ziebach, R., Kahl, B.C., and Wolz, C. (2004) Increased frequency of genomic alterations in *Staphylococcus aureus* during chronic infection is in part due to phage mobilization. *J. Infect. Dis.*, **189**, 724–734.
- 48 Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J.M., Koehrsen, M., Rokas, A., Yandava, C.N., Engels, R., Zeng, E., Olavarrieta, R., Doud, M., Smith, R.S., Montgomery, P., White, J.R., Godfrey, P.A., Kodira, C., Birren, B., Galagan, J.E., and Lory, S. (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl. Acad. Sci. USA*, **105**, 3100–3105.
- 49 Falush, D., Kraft, C., Taylor, N.S., Correa, P., Fox, J.G., Achtman, M., and Suerbaum, S. (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. USA*, **98**, 15056–15061.
- 50 Gressmann, H., Linz, B., Ghai, R., Pleissner, K.P., Schlapbach, R., Yamaoka, Y., Kraft, C., Suerbaum, S., Meyer, T.F., and Achtman, M. (2005) Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.*, **1**, e43.
- 51 Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., and Achtman, M. (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, **445**, 915–918.

- 52 Kraft, C., Stack, A., Josenhans, C., Niehus, E., Dietrich, G., Correa, P., Fox, J.G., Falush, D., and Suerbaum, S. (2006) Genomic changes during chronic *Helicobacter pylori* infection. *J. Bacteriol.*, **188**, 249–254.
- 53 Oh, J.D., Kling-Backhed, H., Giannakis, M., Xu, J., Fulton, R.S., Fulton, L.A., Cordum, H.S., Wang, C., Elliott, G., Edwards, J., Mardis, E.R., Engstrand, L.G., and Gordon, J.I. (2006) The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc. Natl. Acad. Sci. USA*, **103**, 9999–10004.

