ELSEVIER

# Genome evolution in bacteria: order beneath chaos
## Jeffrey G Lawrence and Heather Hendrickson

Bacterial genomes have been viewed as collections of genes, with each gene and genome evolving more-or-less independently through the acquisition of mutational changes. This historical view has been overturned by the finding that genomes of even closely-related taxa differ widely in gene content. Yet, genomes are more than ever-shuffling collections of genes. Some genes within a genome are more transient than others, conferring a layer of phenotypic lability over a core of genotypic stability; this core decreases in size as the taxa included become increasingly diverse. In addition, some lineages no longer experience high rates of gene turnover, and gene content alters primarily through slow rates of gene loss. More importantly, the cell and molecular biology of the bacterial cell imposes constraints on chromosome composition, maintaining a stable architecture in the face of gene turnover. As a result, genomes reflect the sum of processes that introduce variability, which is then arbitrated by processes that maintain stability.

**Addresses**
Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Corresponding author: Lawrence, Jeffrey G (jlawrenc@pitt.edu)

## Introduction
Complete genome sequences have served to define and to describe many genes and regulatory features as well as numerous other elements. They have also revealed the actions of gene gain, loss and rearrangement, and have shed light on the scope and impact of numerous genetic and evolutionary processes such as lateral gene transfer, concerted evolution, and the proliferation of junk DNA. Yet, chromosomes are more than just collections of genes — they are physically large and genetically indispensable polymers that must be replicated, organized, compacted, modified and apportioned to daughter cells in a faithful and timely manner. Perhaps it is only with the age of comparative genomics that appreciation has developed for genomes as a whole — that is, as entities that possess coalescent properties that reflect genome-scale processes.

Here we describe how single genomes are incomplete representatives of the total gene content of a bacterial species, but that this variability varies among lineages. The exchange of genes among genomes is non-random, with some genes being more recalcitrant to transfer or loss than others. Moreover, genes can be delivered *en masse* by way of the acquisition of genomic islands; however, the fate of all genes within these islands remains unclear. This turnover in gene content operates in the context of a genome architecture — that is, sequences that provide structure to replicores — that must be maintained to allow efficient chromosome segregation and partitioning.

## The pan-genome
A genome sequence provides a complete genetic inventory for a particular bacterial strain; however, this strain is only a single representative of a species, the members of which can be genotypically and phenotypically much more diverse [1••]. Although the processes of gene gain and gene loss have been appreciated for some time, the scope of genotypic variation — as measured by gene presence or absence — has been unveiled in several taxa for which multiple complete genomes are available, including (as of May 2005) multiple strains of almost 40 genera that represent 8 bacterial divisions. Most analyses have revealed large differences in gene content between even closely related strains. For example, comparison of the laboratory strain *Escherichia coli* K12 to both uropathogenic and enterohemorrhagic strains revealed that startlingly few genes (<40% of the total number of genes present; Figure 1) were shared by these three strains [2]. One interpretation is that different strains exploit somewhat different ecologies by virtue of metabolic differences imparted by different gene inventories. The collection of genes shared among members of the same species — that is, the clade-specific metagenome — has been termed the 'pan-genome'. Similar levels of variation (Figure 2a) are observed among nine genomes of *E. coli* and *Shigella* (which are essentially strains of *E. coli*) [1••].

Such large-scale gene transfer implies that chromosomes are chaotic collections of genes acquired from numerous sources and that they are little more than holding pens for transient rosters of genetic free-agents. However, further scrutiny of microbial genomes suggests that is not the case. Taxonomically labile genes often encode-functions that affect the cell surface, signal transduction or pathogenicity, whereas genes recalcitrant to transfer

**Figure 1**
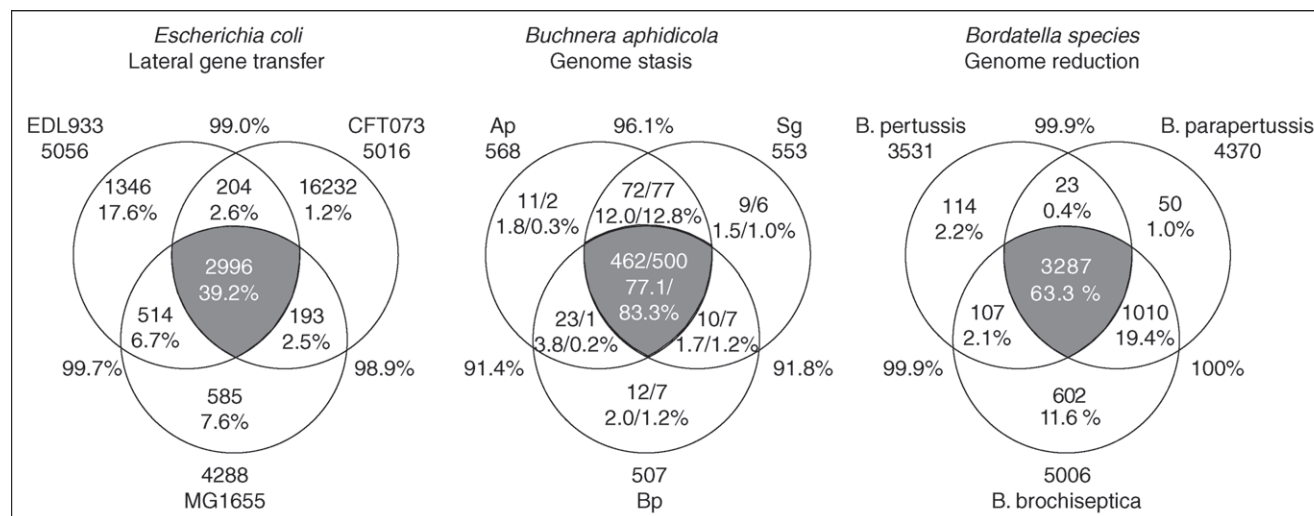


Comparisons of genome content in three sets of Bacteria. Here, the 'pan-genome' is encompassed within the Venn diagram, and the 'core genome' is represented by the shaded region that denotes the genes shared among all three genomes. Numbers inside the Venn diagrams indicate the number of genes (and percentage of total) found to be shared among the indicated genomes; for *Buchnera*, the numbers reflect analyses without/with pseudogenes. Numbers outside the diagram indicate the number of genes in each genome and the percentage identity of the genes that encode the 23S rRNA. Data are presented as reported for genomes of *Escherichia coli* [2], *Buchnera* [15] and *Bordatella* [16]. Lateral gene transfer, large affect of gene acquisition on gene content; genome stasis, little change in gene content; genome reduction, large affect of gene loss on gene content.

are involved in translation or in amino acid biosynthesis [3••]. That is, some genes are transferred more often than others, probably because the functions that they confer are more likely to provide a net advantage to the recipient cell. Therefore, a pan-genome is a variable collection of genes overlaid on a more stable 'core'. Thus, although bacterial genome inventories appear to change dramatically by lateral gene transfer, which allows for rapid adaptation to novel ecological niches, such transfer does not affect all genes equally [3••]. This could preserve the ability to make inferences about the relationships among conserved broadly distributed genes [4].

By contrast, eukaryotic genomics provides a substantially different view of gene acquisition, in which intra-lineage duplication and divergence are much more common than inter-lineage gene transfer. More dramatically, whole-genome duplication is now well supported in the evolution of hemiascomycete yeasts [5,6], and it is strongly suggested that there are multiple duplications in chordates [7,8]. Gene duplications are not thought to play major roles in the expansion of genomic repertoires in bacteria. For example, Lerat *et al.* [9•] examined 13 complete γ-proteobacterial genome sequences and found that variation in genome composition could be attributed to lateral gene transfer; only rarely did expansion of a gene family occur by apparent within-lineage duplication and divergence.

## The core genome
Although some genes may be more recalcitrant to transfer, genes within core genomes might have been transferred or even replaced. Orthologous replacement can introduce new versions of existing genes into genomes. Such transfers can replace even highly conserved genes; for example, lateral transfer of 16S rRNA genes has been described in both Bacteria [10,11] and Archaea [12]. Here, the advantages provided by newly acquired genes are less obvious. In addition, so-called core genes might be replaced by non-homologous counterparts; as the phylogenetic breadth of the genomes examined increases, the number of genes shared among taxa decreased. For example, Charlebois and Doolittle [13•] estimated that Bacteria share between 100 and 150 genes (Figure 2a), whereas only 30–50 genes appear to be shared amongst all free-living prokaryotes — much fewer than required for survival of a free-living cell. This suggests that non-orthologous replacement might be commonplace. More importantly, they established that, as predicted, the core genome becomes smaller when taxa of increasing diversity are examined. The bacterial core of ∼125 genes (Figure 2b) is much smaller than the proteobacterial core of ∼1500 genes or the *E. coli* core of ∼3000 genes (Figure 2a). Thus, although gene turnover has cumulative effects over long periods of time, relatively closely-related genomes might still share large numbers of genes that have potentially consistent relationships [4]. Even so, teasing a consistent phylogenetic

**Figure 2**
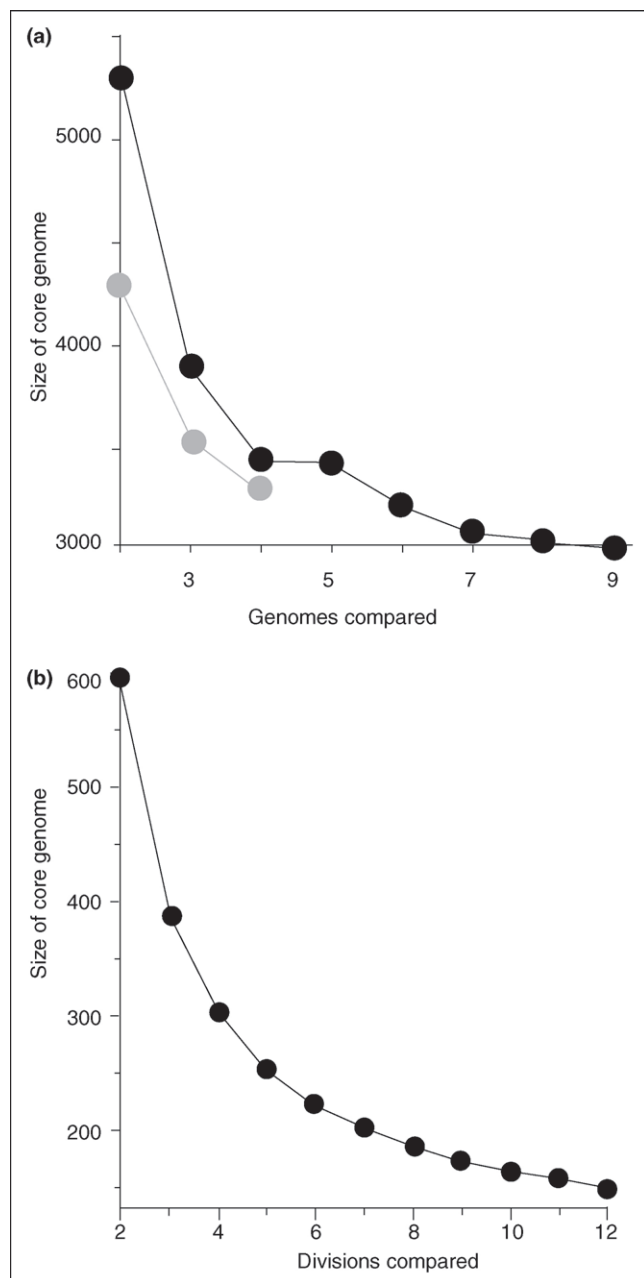


An estimation of the size of 'core genomes'. **(a)** The number of genes are shown that are shared among multiple strains of *E. coli* and *Shigella*; this comprises a core genome. The filled circles represent maximum number of genes shared among the strains listed as reported in [1[••]]; grey circles indicate the average number of genes shared among multiple 2-, 3- and 4-way comparisons. **(b)** Estimation of the size of the Bacterial (non-Archaeal) core. This was predicted from the average number of genes shared between genomes that have undergone an increasing number of bacterial divisions. Data from Charlebois and Doolittle [13[•]].

signal from shared genes is complex, and often one cannot exclude the possibility that core genes have been transferred, or that too little data exist to address this question [14[•]].

Large-scale gene acquisition is not the dominant force that shapes all microbial lineages. For example, comparison of three lineages of the insect endosymbiont *Buchnera* — which are substantially more divergent than strains of *E. coli* — shows essentially that no gene acquisition has occurred over a period of 50–200 million years [15], and that the minor variation in gene content (Figure 1) is caused by low rates of gene loss and formation of pseudogenes, with no gene acquisition being detected in these lineages. Similarly, the genomes of pathogens *Bordatella pertussis* and *Bordatella parapertussis* differ from that of the less virulent *Bordatella bronchiseptica*, primarily owing to gene loss [16]; this is reflected in few genes being unique to either the *B. pertussis* or *B. parapertussis* genome, and many genes lost from the *B. pertussis* genome (Figure 1). Therefore, the pan-genome of some lineages might be closely approximated by the sequences of one or a few genomes. For example, near complete genomes of *Leptospirillum* and *Ferroplasma* were deduced from environmental samples found in an acid hot-spring [17[•]], which suggests that the pan-genomes of these specialized ecologically-restricted organisms are smaller than those of bacteria that adopt a more generalist lifestyle.

The view that phenotypic variation is derived primarily from gene gain and gene loss is easy to adopt as these types of genotypic variation are readily identified and can often be correlated to large-scale changes in function or phenotype (e.g. pathogenicity islands). However, analyses of more subtle differences can highlight function alteration caused by the modification of existing genes — another conventional paradigm for evolutionary change. For example, modification of toxin gene expression by increased promoter efficiency in *B. pertussis* contributes to the pathogenicity of this strain [16]; such modifications accompany massive gene loss rather than gene gain (Figure 1). Similarly, the PmrD protein shares only ~55% identity between *E. coli* and *Salmonella*, which is much lower than for other proteins. Therefore, in *Salmonella*, PmrD-mediated activation of genes in response to low concentrations of $Mg^{2+}$ occurs by virtue of interactions between the PmrD regulatory protein and the PmrAB two-component regulatory system; the *E. coli* PmrD protein fails to interact with PmrAB [18].

## Genomic islands and the flexible genome

The observation that core genomes are smaller when more distantly related strains are included [19] indicates that the story is more complex than the simple classification of genes belonging to either the core genome or the 'flexible' genome. That is, genes that are considered to belong to the core when closely related genomes are compared will be classified as flexible when genomes of more distantly related organisms are compared. Several factors control the non-random patterns of gene exchange among genomes. Beyond their unequal propensities for transfer [3[••]], genes are not mobilized at random. In many

cases, large regions of DNA are integrated — these are denoted as genomic islands [20], a general term that encompasses pathogenicity, symbiosis, ecological and saprophytic islands [19]. Genomic islands are typically large (tens of kilobases), are often associated with tRNAs and a linked integrase gene, and show an unusual nucleotide composition — all of which are signs of foreign origin. They are often identified through bioinformatic analyses of single genomes [21,22], although multiple genome comparison can be helpful.

Although it might be tempting to conclude that genomic islands are adaptive, their large size, retention of genes that encode integrase and other functions probably not selected in the host, and strong atypical character indicate that they have recently arrived in the genome. Over time, all of these features would be lost, so genomic islands probably provide little, if any, benefit to the genome. Genomic islands lose genes by deletion and pseudogene formation [23,24]. In addition, islands contain DNA that does not contribute to purported adaptive phenotypes; for example, a 100 kb insertion in *Salmonella* dedicates only 5 kb to sucrose degradation [25], the proposed selection for maintenance of this island [20]. Similarly, genes in the SPI-3 cluster can be eliminated without having any effect on pathogenicity [24]. Therefore, additional unidentified beneficial functions must be encoded in these islands to prevent the eventual loss of these genes.

If genomic islands are recent acquisitions, what is their role in genome evolution? Like plasmids and prophages, they might provide a method for persistence and a platform for the introduction of many genes in a single event. Although not all genes they introduce will be retained, it remains to be seen how many foreign genes interspersed in the genome arose from former genomic islands. Such large-scale dispersal of genes has been seen in eukaryotes, for example in the dispersal of large numbers of mitochondrial genes in the nuclear genome [26]. That is, genes now considered to be part of a core might have arrived in a genome as part of a genomic island. In addition, genes do not need to remain within a genomic island to be retained.

## More than beads on a string
Although the prevalence of gene exchange could portray genome evolution as little more than the shuffling of the gene repertoires of an organism, further scrutiny has offset this view. That is, there is order and structure to genomes that are maintained in the face of rapid changes in gene content. For example, genes are not distributed randomly between leading and lagging strands. Although the conventional view has been that the higher number of genes encoded on leading strands reflects selection against collisions of RNA and DNA polymerases, Rocha and Danchin [27] have demonstrated that it is the impor-

tance of a gene, not merely its level of expression, which dictates preferential positioning on the leading strand. Aside from gene organization, each replicore (the region from the replication origin to terminus) varies in predictable ways along gradients from the replication origin to the terminus, including origin-proximal placement of important genes [28] and changes in mutational bias towards the replication terminus [29].

Evidence for chromosome organization is not limited to the results of bioinformatic analyses. Advances in both molecular genetics and bacterial cell biology have allowed unprecedented insight into chromosome structure and into the dynamics of DNA movement during the bacterial cell cycle. The bacterial chromosome has long been known to be organized into domains that are defined by their supercoiling character. Transcription imposes a barrier to the diffusion of supercoils [30], which affects processes such as the expression of proximal genes [31] and the movement of mobile elements [32]. Supercoiling domains might affect the degree to which different portions of the *Salmonella* chromosome interact during homologous recombination events [33•], thereby directly influencing rates of chromosome rearrangement. Although the boundaries of supercoiling domains (of which there might be hundreds [34]) are not rigidly defined, it has been demonstrated that supercoiling density is under selection [35]. As a result, sequences that control supercoiling density might constrain chromosome evolution in ways not predicted from gene inventories.

Aside from regional organization, the position and movement of the chromosome itself is highly regulated. Green fluorescent protein fusions to transcriptional repressors have allowed for the fluorescent tagging of specific chromosomal locations, which has enabled direct visualization of chromosome movement. These techniques have allowed the movement of the origin of replication to be tracked towards the poles following replication in both *E. coli* and *Bacillus subtilus*. In *E. coli*, the origin of replication is not the site that is the first to move towards the cell poles; the 25 bp *migS* sequence, which is located close to but not at the origin of replication, directs chromosome movement [36]. The *migS* sequence can function as a 'centromere' of sorts, even when placed far from the origin of replication [37]. In *Vibrio cholerae*, the origins of replication of the two chromosomes were shown to move towards the cell poles at different time points in the cell cycle, which suggests that the factors that regulate the initiation of chromosome movement act differentially on these replicons [38].

In *E. coli*, the termini of replication remain at the cell center following replication, and after septation they move to the center of their respective daughter cells [39]. However, after replication, both termini are often

located on one side of the division septum. If this occurs, one terminus region (importantly, the one 'belonging' to the other daughter cell) is actively translocated across the septum [39]; movement of DNA through a confined passage is visualized as a loss of the bound green fluorescent protein–repressor on one side of the septum and the reappearance of the fluorescent focus on the other side as the GFP protein rebinds to the operator site. Translocation of the chromosome might be mediated by the membrane-bound cell division protein FtsK, which has been observed to translocate along DNA towards the *dif* site [40••] in which the FtsK-delivered XerCD protein acts to resolve dimeric chromosomes [41]. Coordinated movement of other regions of the chromosome have been observed in *Caulobacter*, in which chromosomal loci at defined distances from the replication origin move at predicted times [42••], which suggests that

**Figure 3**



Accumulation of strand-specific sequences at the terminus of replication. **(a)** The abundance of the four octomer permutations of the nonomer 5′GGGGYAGGG on the leading (filled circles) and lagging (open circles) strands of the *E. coli* K12 genome. The distribution of the four octomers is shown for **(b)** the complete *E. coli* genome, as well as **(c)** the genes shared with *Salmonella* and *Klebsiella* and **(d)** the genes unique to *E. coli*. The skewed distribution of these sequences in genes gained by *E. coli* reflects either existing biases in successfully acquired genes or selection to establish this pattern rapidly. Native and acquired genes were identified by Lawrence and Ochman [45]. Vertical lines on the Watson (w) and Crick (c) strands denote the occurrence of an octomer at that position.

genomic architecture is involved in more than merely the positioning of the origin and terminus of replication.

The necessity for the FtsK protein to move towards the replication terminus illustrates the utility of chromosome architecture — that is, directionality or polarity — that cannot be imparted by gene inventory alone. FtsK might recognize sequences distributed preferentially on the leading strand, thereby allowing it to translocate towards the replication terminus; when at this terminus, DNA polarity changes from the leading strand to the lagging strand. Candidate sequences have been identified in many genomes that not only appear to be overrepresented on leading strands [43] but also accumulate towards the terminus of replication (the point at which natural selection for function is the strongest) [44]. For example, the four octomer permutations of 5′GGGGYAGGG in the *E. coli* genome are both biased to the leading strand and increase in abundance towards the terminus of replication (Figure 3a,b). Notably, these octomers are not only biased in genes ancestral to the *E. coli* genome (Figure 3c) but are also biased among genes introduced into the genome by lateral gene transfer (Figure 3d). This bias, even in newly acquired genes, is consistent with octomer composition playing a role in the constraint of gene flow, whereby there is counterselection of strains that acquire gene fragments that severely disrupt patterns of sequences responsible for the orderly replication and segregation of chromosomes [44].

Thus, bacterial genomes in many lineages appear to undergo rapid change through gene acquisition and gene loss. However, beneath this seemingly chaotic exterior lies a more stable chromosomal core, which is conserved both in gene content and in genome architecture. The pattern reflects selection for processes that function at the genomic level — that is, above the level of the gene.
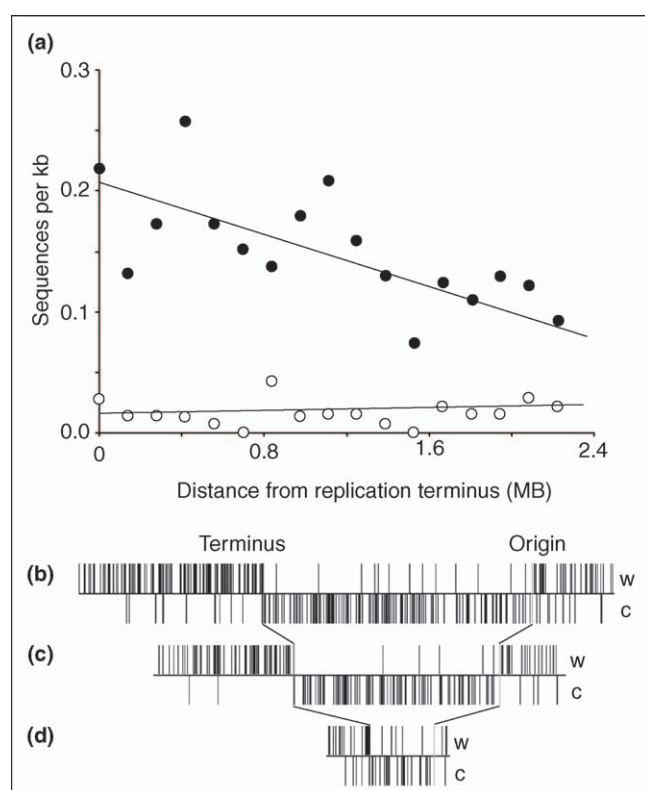
## Conclusions

Several processes affect genome evolution. First, the flow of genes between genomes is influenced both by the variability in efficacy of transfer among different classes of genes, and by the frequent transfer of large blocks of genes as genomic islands. These process result in closely-related strains harboring a 'core' genome augmented by a sampling of genes from its pan-genome. Second, rearrangements within genomes are constrained by sequences that impart genome architecture. Here, chromosome replicores may be structured by the asymmetric distributions of sequences recognized by proteins that act at the replication terminus. As a whole, genes do not evolve in isolation, but rather respond to complex contextual cues and shape genome evolution *en masse*.

## Acknowledgements

# 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. Konstantinidis KT, Tiedje JM: **Genomic insights that advance the**
•• **species definition for prokaryotes**. *Proc Natl Acad Sci USA* 2005, **102**:2567-2572.
The relationship between average nucleotide identity and gene content is explored, showing that genomes that have nearly identical shared genes can vary widely in gene content, whereas genomes with highly similar contents can house more distantly related genes.

2. Welch RA, Burland V, Plunkett G III, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J *et al.*: **Extensive mosaic structure revealed by the complete genome sequence of uropathogenic** *Escherichia coli*. *Proc Natl Acad Sci USA* 2002, **99**:17020-17024.

3. Nakamura Y, Itoh T, Matsuda H, Gojobori T: **Biased biological**
•• **functions of horizontally transferred genes in prokaryotic genomes**. *Nat Genet* 2004, **36**:760-766.
Laterally transferred genes are identified in 116 genomes. These are analyzed for their encoded function; genes that alter the cell surface, genes that contribute to pathogenicity, and regulatory genes are found to be transferred more often than one would expect.

4. Lerat E, Daubin V, Moran NA: **From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria**. *PLoS Biol* 2003, **1**:E19.

5. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S *et al.*: **The** *Ashbya gossypii* **genome as a tool for mapping the ancient** *Saccharomyces cerevisiae* **genome**. *Science* 2004, **304**:304-307.

6. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast** *Saccharomyces cerevisiae*. *Nature* 2004, **428**:617-624.

7. McLysaght A, Hokamp K, Wolfe KH: **Extensive genomic duplication during early chordate evolution**. *Nat Genet* 2002, **31**:200-204.

8. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A *et al.*: **Genome duplication in the teleost fish** *Tetraodon nigroviridis* **reveals the early vertebrate proto-karyotype**. *Nature* 2004, **431**:946-957.

9. Lerat E, Daubin V, Ochman H, Moran NA: **Evolutionary**
• **origins of genomic repertoires in bacteria**. *PLoS Biol* 2005, **3**:e130.
Lateral gene transfer is inferred to change gene repertoires in proteo-bacteria, but many genes appear to show a consistent phylogenetic signal.

10. Yap WH, Zhang Z, Wang Y: **Distinct types of rRNA operons exist in the genome of the Actinomycete** *Thermomonospora chromogena* **and evidence for horizontal transfer of an entire rRNA operon**. *J Bacteriol* 1999, **181**:5201-5209.

11. Eardly BD, Nour SM, van Berkum P, Selander RK: **Rhizobial 16S rRNA and** *dnaK* **genes: mosaicism and the uncertain phylogenetic placement of** *Rhizobium galegae*. *Appl Environ Microbiol* 2005, **71**:1328-1335.

12. Boucher Y, Douady CJ, Sharma AK, Kamekura M, Doolittle WF: **Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes**. *J Bacteriol* 2004, **186**:3980-3990.

13. Charlebois RL, Doolittle WF: **Computing prokaryotic gene**
• **ubiquity: rescuing the core from extinction**. *Genome Res* 2004, **14**:2469-2477.
Genes shared among 147 bacterial genomes are assessed and used to define the 'core' genome at several levels of taxonomic inclusive-ness.

14. Bapteste E, Susko E, Leigh J, MacLeod D, Charlebois RL,
• Doolittle WF: **Do orthologous gene phylogenies really support tree-thinking?** *BMC Evol Biol* 2005, **5**:33.
A clever implementation of heat maps shows that few genes provide strong phylogenetic support for single phylogenies. This confounds the ability to quantify the effects of lateral gene transfer.

15. van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ *et al.*: **Reductive genome evolution in** *Buchnera aphidicola*. *Proc Natl Acad Sci USA* 2003, **100**:581-586.

16. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL *et al.*: **Comparative analysis of the genome sequences of** *Bordetella pertussis*, *Bordetella parapertussis* **and** *Bordetella bronchiseptica*. *Nat Genet* 2003, **35**:32-40.

17. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ,
• Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment**. *Nature* 2004, **428**:37-43.
The metagenome from an acid mine was assessed. This study revealed low levels of both species diversity and genetic diversity within the dominant species.

18. Winfield MD, Groisman EA: **Phenotypic differences between** *Salmonella* **and** *Escherichia coli* **resulting from the disparate regulation of homologous genes**. *Proc Natl Acad Sci USA* 2004, **101**:17162-17167.

19. Hacker J, Carniel E: **Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes**. *EMBO Rep* 2001, **2**:376-381.

20. Dobrindt U, Hochhut B, Hentschel U, Hacker J: **Genomic islands in pathogenic and environmental microorganisms**. *Nat Rev Microbiol* 2004, **2**:414-424.

21. Zhang R, Zhang CT: **Identification of genomic islands in the genome of** *Bacillus cereus* **by comparative analysis with** *Bacillus anthracis*. *Physiol Genomics* 2003, **16**:19-23.

22. Mantri Y, Williams KP: **Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities**. *Nucleic Acids Res* 2004, **32**:D55-D58.

23. Middendorf B, Hochhut B, Leipold K, Dobrindt U, Blum-Oehler G, Hacker J: **Instability of pathogenicity islands in uropathogenic** *Escherichia coli* **536**. *J Bacteriol* 2004, **186**:3086-3096.

24. Blanc-Potard AB, Solomon F, Kayser J, Groisman EA: **The SPI-3 pathogenicity island of** *Salmonella enterica*. *J Bacteriol* 1999, **181**:998-1004.

25. Hochhut B, Jahreis K, Lengeler JW, Schmid K: **CTnscr94, a conjugative transposon found in enterobacteria**. *J Bacteriol* 1997, **179**:2097-2102.

26. Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastiani F, Gelius-Dietrich G, Henze K, Kretschmann E, Richly E, Leister D *et al.*: **A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes**. *Mol Biol Evol* 2004, **21**:1643-1660.

27. Rocha EP, Danchin A: **Essentiality, not expressiveness, drives gene-strand bias in bacteria**. *Nat Genet* 2003, **34**:377-378.

28. Rocha EP: **The replication-related organization of bacterial genomes**. *Microbiology* 2004, **150**:1609-1627.

29. Daubin V, Perriere G: **G+C3 structuring along the genome: a common feature in prokaryotes**. *Mol Biol Evol* 2003, **20**:471-483.

30. Deng S, Stein RA, Higgins NP: **Transcription-induced barriers to supercoil diffusion in the** *Salmonella typhimurium* **chromosome**. *Proc Natl Acad Sci USA* 2004, **101**:3398-3403.

31. Jeong KS, Ahn J, Khodursky AB: **Spatial patterns of transcriptional activity in the chromosome of** *Escherichia coli*. *Genome Biol* 2004, **5**:R86.

32. Manna D, Breier AM, Higgins NP: **Microarray analysis of transposition targets in** *Escherichia coli*: **the impact of transcription**. *Proc Natl Acad Sci USA* 2004, **101**:9780-9785.

33. Garcia-Russell N, Harmon TG, Le TQ, Amaladas NH, Mathewson
• RD, Segall AM: **Unequal access of chromosomal regions to each other in *Salmonella*: probing chromosome structure with phage lambda integrase-mediated long-range rearrangements**. *Mol Microbiol* 2004, **52**:329-344.
Genetic techniques are used to assay the ability of sequences to recombine with one another when placed at different chromosome locations. Some chromosomal loci showed a reduced ability to interact with each other.

34. Stein RA, Deng S, Higgins NP: **Measuring chromosome dynamics on different time scales using resolvases with varying half-lives**. *Mol Microbiol* 2005, **56**:1049-1061.

35. Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D: **Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection**. *Genetics* 2005, **169**:523-532.

36. Yamaichi Y, Niki H: ***migS*, a *cis*-acting site that affects bipolar positioning of *oriC* on the *Escherichia coli* chromosome**. *EMBO J* 2004, **23**:221-233.

37. Fekete RA, Chattoraj DK: **A *cis*-acting sequence involved in chromosome segregation in *Escherichia coli***. *Mol Microbiol* 2005, **55**:175-183.

38. Fogel MA, Waldor MK: **Distinct segregation dynamics of the two *Vibrio cholerae* chromosomes**. *Mol Microbiol* 2005, **55**:125-136.

39. Lau IF, Filipe SR, Soballe B, Okstad OA, Barre FX, Sherratt DJ: **Spatial and temporal organization of replicating *Escherichia coli* chromosomes**. *Mol Microbiol* 2003, **49**:731-743.

40. Pease PJ, Levy O, Cost GJ, Gore J, Ptacin JL, Sherratt D,
•• Bustamante C, Cozzarelli NR: **Sequence-directed DNA translocation by purified FtsK**. *Science* 2005, **307**:586-590.
Single-molecule experiments on purified *E. coli* FtsK protein show that the enzyme translocates on DNA in a sequence-specific fashion towards the terminus of replication, as defined by the *dif* site.

41. Ip SC, Bregu M, Barre FX, Sherratt DJ: **Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination**. *EMBO J* 2003, **22**:6399-6407.

42. Viollier PH, Thanbichler M, McGrath PT, West L, Meewan M,
•• McAdams HH, Shapiro L: **Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication**. *Proc Natl Acad Sci USA* 2004, **101**:9257-9262.
Time-lapse photography shows coordinated post-replication movement of numerous loci spread across the *Caulobacter* chromosome. Sequential movement of loci at comparable speeds supports the idea that the same mechanism that moves the origin of replication coordinates movement of the remainder of the chromosome.

43. Worning P, Jensen LJ, Hallin PF, Staerfeldt H-H, Ussery DW: **Origin of replication in circular prokaryotic chromosomes**. *Environ Microbiol* 2005, in press.

44. Lawrence JG, Hendrickson H: **Chromosome structure and constraints on lateral gene transfer**. *Dev Dyn* 2004, **2004**:319-336.

45. Lawrence JG, Ochman H: **Reconciling the many faces of gene transfer**. *Trends Microbiol* 2002, **10**:1-4.