Horizontal gene transfer and bacterial diversity

CHITRA DUTTA* and ARCHANA PAN

Human Genetics and Genomics Group, Indian Institute of Chemical Biology, 4, Raja SC Mullick Road, Kolkata 700 032, India

**Corresponding author (Fax, 91-33-473 5193; Email, cdutta@iicb.res.in).*

Bacterial genomes are extremely dynamic and mosaic in nature. A substantial amount of genetic information is inserted into or deleted from such genomes through the process of horizontal transfer. Through the introduction of novel physiological traits from distantly related organisms, horizontal gene transfer often causes drastic changes in the ecological and pathogenic character of bacterial species and thereby promotes microbial diversification and speciation. This review discusses how the recent influx of complete chromosomal sequences of various microorganisms has allowed for a quantitative assessment of the scope, rate and impact of horizontally transmitted information on microbial evolution.

[Dutta C and Pan A 2002 Horizontal gene transfer and bacterial diversity; *J. Biosci.* (*Suppl. 1*) **27** 27–33]

Introduction

Within mammalian gut flora or inside a distilled water reservoir; under Antarctic ice or in near-boiling water; in acid springs or in alkaline pools – microbial life exists everywhere. More than 10^9 bacterial species are believed to inhabit the biosphere (Dykhuizen 1998; Lawrence 1999). In view of the fact that bacteria are unicellular organisms with a comparitively small genome size, variations observed in their cellular architectures, metabolic properties and phenotypic traits are remarkable (Casjens 1998). Even within narrow taxonomic groups such as the enteric bacteria, a great deal of phenotypic diversity among species is exhibited. In order to display such diverse ecological and phenotypic characters, bacteria must have extremely dynamic genomes, entertaining rapid acquisition, deletion and rearrangements of relevant genetic information.

Two broad classes of mechanisms may be invoked to explain the overwhelming genetic diversity that graces the bacterial world: (i) internal modifications of genetic information, i.e. variations arising out of the divergence and vertical transmission of existing genes that usually take place through accumulation of mutation (Whittam 1996) and inter-genomic homologous recombination (Milkman 1997) and (ii) acquisition or loss of specific set of genes from other species through the process of horizontal or lateral gene transfer (HGT or LGT) (Syvanen and Kado 1998; Lawrence and Roth 1999). Point mutations may lead to "slow" but continuous modification of existing genes, thereby allowing gradual niche expansion. They might be responsible for diversification and speciation of microorganisms on an evolutionary time-scale. HGT, on the other hand, results in abrupt large-scale alterations in the structure and organization of genomes and is, therefore, capable of generating new variants of bacterial strains by "genetic quantum leaps" (Falkow 1996). There are reasons to believe that drug-resistant strains of *Mycobacterium tuberculosis*, and *Streptococcus pneumoniae*, the new *Vibrio cholerae* serotype O139, enterohaemorrhagic *Escherichia coli* of serotype O157 and the *Haemophilus influenzae* biotype aegypticus have all emerged through macroevolution events mediated through HGT (Achtman and Hakenbeck 1992; Hacker *et al* 1997).

Till recently, most studies on microbial evolution focused primarily on the vertical transmission of genetic information. Little attention has been paid to the importance of HGT in shaping bacterial genomes, probably

Keywords. Amelioration; atypical genes; bacterial genomes; microbial speciation; mobile genetic elements; novel traits

because with limited sequence data, the role of horizontal transfer in bacterial evolution was difficult to study. The recent cavalcade of complete genome sequences provides an opportunity to have an insight into the scope (Aravind *et al* 1998; Huynen and Bork 1998; Snel *et al* 1999; Wolf *et al* 1999a), rate (Lawrence and Ochman 1998; Lawrence and Roth 1999) and impact (Lawrence 1997; Lawrence and Roth 1998) of HGT. There is growing evidence that HGT may occur across vast phylogenetic distances, such as from bacteria to eukaryotes (Doolittle 1998), from animals to bacteria (Wolf *et al* 1999a), from bacteria to archaea (Nelson *et al* 1999) and so on. It appears that some genes have flowed 'randomly' through the biosphere, almost as if all life forms constituted one global organism.

2. What types of genes are susceptible to horizontal transfer?

Though considerable HGT seem to have occurred within and across prokaryotes and eukaryotes, not all genes are equally likely to be transferred. For HGT to be successful the acquired genes must persist in the host chromosome; the gene would persist only if it provided a selective benefit to the recipient organism (Lawrence 1999). Essential genes, ubiquitously present in all organisms, like those encoding rRNA operons, are therefore less likely to undergo successful transfer as genomes naïve to their functions are rare. Horizontal transfer of genes under weak or transient selection, in contrast, could greatly benefit lineages that do not contain any functional orthologs of these genes (Lawrence 1999).

In bacteria, the genes whose products catalyze steps in a single pathway and provide for weakly selected functions, often tend to assemble into operons (even when their products do not interact physically), probably to facilitate horizontal transfer of such genes into naïve genomes (Lawrence and Roth 1996). Upon introgression, an operon provides a novel metabolic function allowing its new host to exploit effectively a novel ecological niche (Lawrence 1997; Lawrence and Roth 1998). The physical proximity of the genes in an operon, however, provides no selective benefit to the individual organism – it rather enhances the fitness of the operon itself by allowing efficient horizontal transfer of its constituent genes, otherwise susceptible to loss by genetic drift. The operon organization is therefore considered to be a 'selfish' property of the constituent genes (Lawrence 1997).

According to the complexity hypothesis (Jain *et al* 1999), genes participating in transcription, translation and related process (informational) genes are seldom transferred because they are, in general, members of large, complex systems, products of which are less likely to function successfully in a foreign cytoplasm. Operational (housekeeping) genes (Rivera *et al* 1998), on the other hand, are typically members of small assemblies of few genes and hence can undergo horizontal transfer more frequently.

However, HGT of genes involved in essential functions cannot be completely ruled out, as exemplified by the horizontal transfer of isoleucyl-tRNA synthetase into the *Mycobacterium tuberculosis* genome (Sassanfar *et al* 1996). The close similarity of the *rrnE* region of *Salmonella* subspecies I to that of *E. coli* suggests lateral transfer of this locus between these taxa (Perez Luz *et al* 1998). Recently, a detail analysis of rRNA genes in *Thermonospora chromogena* has revealed the existence of two distinct sets of functional, expressed rRNA genes in the genome, one of which was probably introduced by HGT (Yap *et al* 1999). Experimental evidence suggests that complex pathways can also undergo successful HGT, as reported in case of a cytochrome C biogenesis and export pathway (Kranz and Goldman 1998).

3. How can one identify transferred genes?

How can one establish whether a new trait or a specific region of a genome is the result of a horizontal transfer? Obviously, it may not always be possible to have direct experimental evidence for conversion of a deficient strain in the presence of an appropriate donor. But genes acquired through HGT can often be recognized by their distinctive phenotypic properties or atypical sequence characteristics and restricted phylogenetic distribution in specific lineages.

Each transfer event introduces a specific set of genes into a single lineage. As a result, the acquired trait remains restricted to the descendents of the recipient strain only and is absent from closely related taxa, thereby producing a scattered phylogenetic distribution (Ochman *et al* 2000). In some cases, it is possible to establish the evolutionary history of a gene by analysing its distribution among various lineages. If a gene is confined to one taxon or species, it is more likely to have been acquired through gene transfer than to have been lost independently from multiple lineages. However, one cannot rule out the possibility that a particular phenotypic trait such as resistance to certain antibiotics has evolved independently in diverse lineages through point mutations in existing genes (Ochman *et al* 2000). Hence, it may not always be possible to distinguish between convergent evolution and horizontal transfer on the basis of phylogenetic analyses alone.

In fact, the best clues to the origin and ancestry of any gene within a bacterial chromosome are usually provided by the intrinsic sequence characteristics of the gene itself.

In any chromosome, ancestral (vertically transmitted) genes experience a particular set of directional mutation pressures (Sueoka 1988), mediated by the specific features of the replication machinery of the cell, such as the balance of the dNTP pools, mutational biases of the DNA polymerases, efficiency of mismatch repair systems and so on (Lawrence 1999). Apart from the mutational bias, there are several other characteristic features such as synonymous codon bias (Ikemura 1985; Sharp and Li 1987; Andersson and Kurland 1990; Pan *et al* 1998), fractal distribution of nucleotides (Jeffrey 1990; Dutta and Das 1992) or dinucleotide relative abundance (Karlin *et al* 1997) which leave distinct 'fingerprints' on sequences native to that cytoplasm, while the 'foreign' genes, i.e. genes acquired through HGT, retain the characteristics of the donor genome and thus can be distinguished from ancestral DNA (Lawrence and Ochman 1998). Comparative analyses of *E. coli* and *Salmonella enterica* chromosomes have revealed that a large number of *S. enterica* genes, which are not present in *E. coli* (and other closely related enteric species), have nucleotide and codon compositions significantly different from the characteristic G + C-content and codon bias of the rest of the chromosome (Groisman *et al* 1992; Lawrence and Ochman 1997). Regions adjacent to genes identified as being laterally acquired ones often contain relics of sequences that might have helped in their integration, such as remnants of translocation elements, attachment sites of phage integrases, transfer origins of plasmids, which further affirm their recent integration in the genome.

Based on these concepts, various methods have been developed for identification of potential foreign genes. Among them, are the direct methods such as subtractive hybridization (Lan and Reeves 1996) as well as indirect approaches like assessment and comparison of overall G + C-content (Groisman *et al* 1992), codon-position-specific nucleotide composition (Lawrence and Ochman 1997, 1998), codon usage pattern (Karlin *et al* 1998a,b), dinucleotide relative abundance signature (Karlin and Burge 1995) and Markov Chain Analyses of oligonucleotide biases (Hayes and Borodovsky 1998).

Atypical sequence characteristics however, can also be exhibited by genes whose products perform some special functions and therefore have a biased amino acid composition. In order to find out the total number of putative horizontally transferred genes in a genome, appropriate corrections should be made for such native genes that have selectively evolved to atypical compositions. In other words, the genes for which nucleotide sequence characteristics depart from the prevalent features of their resident genome, but where amino acid compositions of the corresponding gene-products are not unusually biased, have a strong potential of being horizontally acquired.

4. What are the extent and rate of transfer in bacteria?

Availability of complete genome sequences of diverse bacterial species and the ability to recognize horizontally transferred genes solely on the basis of their intrinsic sequence characteristics allows one to estimate the total amount of putative foreign genes in such genomes without resorting to database searching or phylogenetic analyses. Analysis of 19 complete genomes (Ochman *et al* 2000) has revealed that the amount of laterally acquired sequences vary widely among bacteria – ranging from virtually none in certain bacteria with small genome sizes such as *Rickettsia prowazekii*, *Borrelia burgdoferi* and *Mycoplasm genitalium*, to nearly 17% in *Synechocystis* PCC6803 and 18% in *E. coli*. It is likely that the intracellular habitat of some bacterial organisms such as *R. prowazekii* and *M. genitalium* shields these organisms from exposure to potential gene donors, thereby reducing the possibility of acquiring foreign sequences. One should, however, remember that in all these cases the extent of HGT might have been underestimated, for the existing intrinsic methods, though quite efficient in detecting genes acquired in the recent past still fail to recognize very ancient horizontal events, such as those disseminating transfer RNA synthetases (Wolf *et al* 1999b).

To assess the rate of effective horizontal transfer in any organism, one must estimate the amount of foreign DNA segments persisting in the host chromosome and maintaining their protein-coding potential (maintenance of coding potential implies that nonsense mutations in these genes have been eliminated by natural selection). The persistence of a foreign gene in a bacterial genome can be estimated by exploring the patterns of nucleotide substitutions that have taken place since its arrival (Lawrence 1999). At the time of introduction, laterally transferred genes bear the sequence characteristics of the donor genome. But once acquired, they come under the influence of directional mutation pressure appropriate to the recipient genome and eventually start to reflect the DNA composition of their new host. This process of 'amelioration' – whereby a horizontally acquired gene adjusts itself to the base composition and codon bias of the vertically inherited genes of the recipient chromosome – is a function of the characteristic mutational bias of the first, second and third codon positions of the later genes. By estimating the degree to which foreign genes have ameliorated to resemble native genes, Lawrence and Ochman (1997, 1998) determined the average rate of effective horizontal gene transfer in *E. coli* to be 16 kb/Myr, resulting in the introduction of ~ 1600 kb of DNA into the *E. coli* genome since its divergence from the *Salmonella* lineage. Although most of the acquired genes were subsequently deleted, 755 ORFs (547·8 kb)

representing at least 234 acquisition events have persisted (18% of the current chromosome) (Martin 1999). It is difficult to identify fully ameliorated genes that have already attained the characteristic sequence features of the recipient *E. coli* chromosome.

5. How are the genes transferred?

For effective transfer of genetic material across species, three steps need to occur successfully: (i) delivery of the DNA sequence from the donor into the recipient cell; (ii) incorporation of the acquired sequence into the genome of the recipient (or, into an autonomous replicating element such as a plasmid) and (iii) expression of the acquired gene(s) at a significant level in the new environment. While the third step depends critically on the compatibility of the transferred genes with the transcriptional and translational machinery of the host organism, the first two steps are largely indiscriminate with respect to the functional features of acquired sequences.

Bacteria exploit any of the three principal mechanisms for interspecies transfer of genetic elements:

(i) *Transformation*: Through this mechanism, naked DNA can be transmitted between two organisms, even distinctly related ones, provided the donor and recipient cells are present at the same place at the same time. Some bacterial species such as *Bacillus subtilus* and *Streptococcus pnemoniae* can exhibit high levels of transformation only upon reaching specific physiological stages in their life-cycles (Dubnau 1999). But species like *Neisseria gonnhorae* and *Haemophillus influenzae* are perpetually competent to take up foreign DNA, as their transformation proficiency is enhanced by the presence of specific recognition sequences (5′-GCCGTCTGAA-3′ and 5′-AAGTGCGGT-3′ respectively) at significant high frequencies in their respective genomes (Smith *et al* 1985; Goodman and Scocca 1988; Elkins *et al* 1991).

(ii) *Transduction*: This involves bacteriophage-mediated transfer of genetic materials between organisms recognized by the phage (Schicklmaier and Schmieger 1995; Jiang and Paul 1998). In this method, phage-encoded proteins can promote the integration of the transferred sequence into the host chromosome to protect it from degradation by host restriction endonucleases.

(iii) *Conjugation*: In this mechanism, transmission of DNA from a donor to a recipient strain can be mediated by a plasmid, which is self-transmissible or mobilizable or which can integrate into the chromosome forming an Hfr strain (Buchanan-Wollaston *et al* 1987; Heinemann and Sprague 1989). Conjugation may also take place through conjugative transposons. These transposons encode proteins required for their excision form the donor chromosome, formation of a conjugative bridge and transposition into the recipient strain. By conjugation, genetic materials can be exchanged even between different domains (between bacteria and plants, between bacteria and yeast or even between bacteria and mammalian cells) (Buchanan-Wollaston *et al* 1987; Heinemann and Sprague 1989; Grillot-Courvalin *et al* 1998) provided the donor and recipient cells are in physical contact with one another.

Insertion of donor DNA into a recipient cytoplasm does not ensure successful gene transfer unless the newly entered sequence remains stable in the host chromosome. Though the species of *Haemophillus* are characterized by natural competence in transformation, the number of horizontally transferred genes in *H. influenzae Rd* genome is quite low compared to other bacterial chromosomes (Ochman *et al* 2000). Stable incorporation of foreign DNA into bacterial genomes can be mediated by any of the following processes: (i) homologous recombination (usually takes place among closely related taxa), (ii) persistence as an episome (if favoured by natural selection), (iii) integration mediated by mobile genetic elements and (iv) illegitimate incorporation through chance double strand break repair.

Although these mechanisms enable bacteria to acquire sequences from diverse organisms, several processes such as restriction modification systems, mismatch repair systems, or the differential recognition by DNA uptake systems exist to limit the integration of foreign genetic elements into bacterial genomes.

6. Why is gene transfer so frequent?

Horizontal transfer of genes appears to be a common practice between microbes, probably due to the important role of such transfer in shaping the architecture of microbial genomes, conferring novel metabolic capabilities to the recipient genome and enabling an organism to explore new ecological niches. The impact of interspecies gene transfer is radically different from that of the spontaneous mutation. Point mutations can lead only to the subtle refinement and alteration of existing metabolic functions, but horizontal gene transfer has the capability of introducing, immediately upon integration, completely novel physiological traits as discussed below:

(i) *Antibiotic resistance*: Bacterial organisms may acquire antibiotic resistance genes through horizontal transfer (de la Cruz 2000). Microbial genome analyses have shown that the resistance determinants in such organisms are often associated with mobile genetic elements like plasmids, integrons or complex transposons that promote their transfer between bacterial genomes (Grinsted *et al* 1990; Freiberg *et al* 1997; Mazel *et al* 1998). For example, a three-gene operon conferring resistance to kanamycin,

bleomycin and streptomycin is flanked by two IS50 insertion sequence in the transposon Tn5 (Berg 1989), which can integrate into the chromosomes of physiologically diverse bacterial species.

(ii) *Pathogenicity islands*: The sporadic distribution of pathogenic organisms has long suggested that bacterial virulence occurs not by mutation, but by horizontal acquisition of pathogenicity determinants. Discovery of large virulence plasmids in pathogenic *Shigella* and *Yersinia* (Gemski *et al* 1980; Portnoy *et al* 1981; Maurelli *et al* 1985; Sasakawa *et al* 1988) or the conversion of laboratory strains of *E. coli* from avirulent to virulent forms upon experimental introduction of genes from other species (Isberg and Falkow 1985; McDaniel and Kaper 1997) support the notion of horizontal transfer of virulence factors in bacteria.

Recent genomic studies have revealed that virulence genes often exist in the bacterial chromosome as discrete gene clusters, usually referred to as "virulence cassettes" or "pathogenicity islands" (Groisman and Ochman 1996; Hacker *et al* 1997). These pathogenicity islands often reside at tRNA and tRNA-like loci, which appear to be common sites for integration of foreign sequences (Hacker *et al* 1997; Ochman *et al* 2000). The sequences flanking such islands frequently include short direct repeats, reminiscent of those generated upon integration of mobile genetic elements, while ORFs in certain pathogenicity islands exhibit sequence similarity to bacteriophage integrases (Ochman *et al* 2000). These observations suggest that horizontal acquisition of virulence gene clusters is a common strategy of bacterial organisms for undergoing transformation from the benign form into a pathogen.

(iii) *Metabolic traits*: In bacteria, many species-specific metabolic traits can be ascribed to HGT. The host chromosome usually retains an acquired gene when its product confers some useful metabolic capabilities. For example, horizontal acquisition of the *lac* operon has enabled *E. coli* to use the milk sugar lactose a carbon source and thereby, to explore a new niche like the mammalian colon (Ochman *et al* 2000).

Microbial genome analyses have depicted scattered distribution of some specific sets of genes in phylogenetically divergent organisms that have independently adapted to a common life-style, such as archeal and bacterial hyperthermophiles or the intracellular pathogens *Rickettsia* and *Chlamydia* (Wolf *et al* 1999a). It is tempting to speculate that these genes of unknown functions have been acquired by these organisms through horizontal transfer, in order to meet the metabolic and physiological requirements of these special environments.

Comparative analysis of the complete genome sequences of *E. coli* and *S. enterica* has revealed that none of the phenotype traits that distinguish the two species are attributable to the point mutational change of homologous genes. Instead, all of the species-specific traits derive from functions encoded either by horizontally acquired genes (e.g. lactose utilization, citrate utilization, propanediol utilization, indole production) or from the loss of ancestral genes (e.g. alkaline phosphate). It is therefore evident, that horizontal transfer of genes play an important role in providing novel metabolic capabilities and thereby in catalyzing the diversification of bacterial lineages.

7. Concluding remarks

As more and more genome sequences are determined, it is becoming clear that inter-species transmission of genetic information is pervasive among microrganisms; that it may occur at vast phylogenetic distances and that it can change the ecological and pathogenic character of bacterial species. Basic concepts of biological evolution therefore, need to be reviewed (Doolittle 1999). A single universal phylogenetic tree might not be the best way to depict relationships between all living and extinct species. Instead, a web- or net-like pattern might provide a more appropriate representation (Doolittle 1999). As Sonea and Paniset (1976) and Reanney (1978) suggested several decades ago all prokaryotes together might be considered as one 'global superorganism' divided into sub-populations within and between which genes are exchanged at varying frequencies.

Rampant inter-species gene transfer has also redefined the concept of the 'universal ancestor' (Woese 1998). There is no guarantee that a gene currently found in all three domains of life, Bacteria, Archea and Eukarya, was present in their common ancestor – it could have arisen more recently in one domain and spread to the others. As stated by Woese (1998): 'the universal ancestor is not a discrete entity. It is, rather, a diverse community of cells that survives and evolves as a biological unit.'

Since bacterial genomes are not growing ever larger in size, the acquisition of foreign genetic elements must be counter-balanced by the loss of native genes (Lawrence 1999, Lawrence and Roth 1999). Analysis of genome sequences of various host-dependent bacteria, such as *Mycoplasma*, *Chlamydia* and *Rickettsia*, show that bacterial genomes have a tendency to lose non-essential DNA (Andersson and Kurland 1998; Andersson and Andersson 1999). Acquired genes that confer metabolic traits necessary for niche expansion may be maintained, while genes providing little selective benefit may be lost. In some cases, loss of function may be selectively advantageous, as exemplified by the beneficial loss of lysine decarboxylase by *Shigella* and enteroinvasive *E. coli* (Maurelli *et al* 1998). It is, therefore, not gene acquisition *per se*, but an optimization between acquisition and loss of genes

that redefines the ecological niche of a microorganism, which in effect, promotes speciation.

Acknowledgements

AP is grateful to the Council of Scientific and Industrial Research, New Delhi, for a Research Associateship.

References

Achtman M and Hakenbeck R 1992 Recent developments regarding the evolution of pathogenic bacteria; in *Molecular biology of bacterial infection: Current status and future perspective* (ed.) C E Hormaeche (Cambridge: SGM/Cambridge University) pp 13–31

Andersson J O and Andersson S G E 1999 Insights into the evolutionary process of genome degradation; *Curr. Opin. Genet. Dev.* **9** 664–671

Andersson S G E and Kurland C G 1990 Codon preferences in free-living microorganisms; *Microbiol. Rev.* **54** 198–210

Andersson S G E and Kurland C G 1998 Reductive evolution of resident genomes; *Trends Microbiol.* **6** 263–268

Aravind L, Tatusov R L, Wolf Y I, Walker D R and Koonin E V 1998 Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles; *Trends Genet.* **14** 442–444

Berg D E 1989 in *Mobile DNA* (eds) D E Berg and M M Howe (Washington DC: American Society for Microbiology) pp 185–210

Buchanan-Wollaston V, Passiatore J E and Canon F 1987 The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants; *Nature* (*London*) **328** 170–175

Casjens S 1998 The diverse and dynamic structure of bacterial genomes; *Annu. Rev. Genet.* **32** 339–377

de la Cruz F 2000 Horizontal gene transfer and the origin of species: lessons from bacteria; *Trends Microbiol.* **8** 128–133

Doolittle W F 1998 You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes; *Trends Genet.* **14** 307–311

Doolittle W F 1999 Lateral genomics; *Trends Cell Biol.* **9** M5-M8

Dubnau D 1999 DNA uptake in bacteria; *Annu. Rev. Microbiol.* **53** 217–244

Dutta C and Das J 1992 Mathematical characterization of chaos game representation: New algorithms for nucleotide sequence analysis; *J. Mol. Biol.* **228** 715–729

Dykhuizen D E 1998 Santa Rosalia revisited: why are there so many species of bacteria?; *Antoine Van Leeuwenhoek J. Microbiol. Serol.* **73** 25–33

Elkins C, Thomas C E, Seifert H S and Sparling P F 1991 Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence; *J. Bacteriol.* **173** 3911–3913

Falkow S 1996 The evolution of pathogenicity in *Escherichia*, *Shigell*a and *Salmonella*; in *Escherichia* coli and *Salmonella: Cellular and Molecular Biology* (ed.) F C Neidhardz (Washington DC: American Society for Microbiology) pp 2723–2729

Freiberg C, Fellay R, Bairoch A, Broughton W J and Rosenthal A 1997 Molecular basis of symbiosis between *Rhizobium* and legumes; *Nature* (*London*) **387** 394–401

Gemski P, Lazere J R, Casey T and Wohlhieter J A 1980 Presence of virulence-associated plasmid in *Yersinia pseudotuberculosis*; *Infect. Immun.* **28** 1044–1047

Goodman S D and Scocca J J 1998 Identification and arrangement of DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*; *Proc. Natl. Acad. Sci. USA* **85** 6982–6986

Grillot-Courvalin C, Goussard S, Huetz F, Ojcius D M and Courvalin P 1998 Functional gene transfer from intracellular bacteria to mammalian cells; *Nat. Biotechnol.* **16** 862–866

Grinsted J, de la Cruz F and Schmitt R 1990 The Tn21 subgroup of bacterial transposable elements; *Plasmid* **24** 163–189

Groisman E A and Ochman H 1996 Pathogenicity islands: bacterial evolution in quantum leaps; *Cell* **87** 791–794

Groisman E A, Saier M H Jr and Ochman H 1992 Horizontal transfer of a phosphatase gene as evidence for the mosaic structure of the *Salmonella* genome; *EMBO J.* **11** 1309–1316

Hacker J, Blum-Oehler G, Mühldorfer I and Tschäpe H 1997 Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution; *Mol. Microbiol.* **23** 1089–1097

Hayes W S and Borodovsky M 1998 How to interpret an anonymous bacterial gonome: machine learning approach to gene identification; *Genome Res.* **8** 1154–1171

Heinemann J A and Sprague G F J 1989 Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast; *Nature* (*London*) **340** 205–209

Huynen M A and Bork P 1998 Measuring genome evolution; *Proc. Natl. Acad. Sci. USA* **95** 442–444

Ikemura T 1985 Codon usage and tRNA content in unicellular and multicellular organisms; *Mol. Biol. Evol.* **2** 13–34

Isberg R R and Falkow S A 1985 A single genetic locus encoded by *Yersima pseudotuberculosis* permits invasion of cultured animal cells by *Escherichia coli* K-12; *Nature* (*London*) **317** 19–25

Jain R, Rivera M C and Lake J A 1999 Horizontal gene transfer among genomes: the complexity hypothesis; *Proc. Natl. Acad. Sci. USA* **96** 3801–3806

Jeffrey H J 1990 Chaos game representation of gene structure; *Nucleic Acids Res.* **18** 2163–2170

Jiang S C and Paul J H 1998 Gene transfer by transduction in the marine environment; *Appl. Environ. Microbiol.* **64** 2780–2787

Karlin S and Burge C 1995 Dinucleotide relative abundance extremes: a gonomic signature; *Trends Genet.* **11** 283–290

Karlin S, Campbell A M and Mrázek J 1998a Comparative DNA analysis across diverse genomes; *Annu. Rev. Genet.* **32** 185–225

Karlin S, Mrázek J and Campbell A M 1998b Codon usages in different gene classes of the *Escherichia coli* genome; *Mol. Microbiol.* **29** 1341–1355

Karlin S, Mrazek J and Cambell A 1997 Compositional biases of bacterial genomes and evolutionary implications; *J. Bacteriol.* **179** 3899–3913

Kranz R G and Goldman B S 1998 Evolution and horizontal transfer of an entire biosynthetic pathway for cytochrome biogenesis: *Helicobacter*, *Deinococcus*, Archae and more; *Mol. Microbiol.* **27** 871–874

Lan R and Reeves P 1996 Gene transfer is a major force in bacterial evolution; *Mol. Biol. Evol.* **13** 47–55

Lawrence J G 1997 Selfish operons and speciation by gene transfer; *Trends Microbiol.* **5** 355–359

Lawrence J G 1999 Gene transfer, speciation, and the evolution of bacterial genomes; *Curr. Opin. Microbiol.* **2** 519–523

Lawrence J G and Ochman H 1997 Amelioration of bacterial genomes: rates of change and exchange; *J. Mol. Evol.* **44** 383–397

Lawrence J G and Ochman H 1998 Molecular archaeology of the *Escherichia coli* genome; *Proc. Natl. Acad. Sci. USA* **95** 9413–9417

Lawrence J G and Roth J R 1996 Selfish operons: horizontal transfer may drive the evolution of gene clusters; *Genetics* **143** 1843–1860

Lawrence J G and Roth J R 1998 Roles of horizontal transfer in bacterial evolution; in *Horizontal transfer* (eds) M Syvanen and C I Kado (London: Chapman and Hall) 208–225

Lawrence J G and Roth J R 1999 Genomic flux: genome evolution by gene loss and acquisition; in *Organization of the prokaryotic genome* (ed.) R L Charlebois (Washington DC: American Society for Microbiology) pp 263–289

Martin W 1999 Mosaic bacterial chromosomes: a challenge en root to a tree of genomes; *Bioessays* **21** 99–104

Maurelli AT, Baudry B, d'Hauteville H, Hale T L and Sansonetti P J 1985 Cloning of plasmid DNA sequences involved in invasion of HeLa cells by *Shigella flexneri*; *Infect. Immun.* **49** 164–171

Maurelli A T, Fernández R E, Bloch C A, Rode C K and Fasano A 1998 '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* **95** 3943–3948

Mazel D, Dychinco B, Webb V A and Davies J 1998 A distinctive class of integron in the *Vibrio Cholerae* genome; **280** 605–608

McDaniel T K and Kaper J B 1997 A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12; *Mol. Microbiol.* **23** 399–407

Milkman R 1997 Recombination and population structure in *Escherichia coli*; *Genetics* **146** 745–750

Nelson K E *et al* 1999 Evidence for lateral gene transfer between Archaea and Bacteria from the genome sequence of *Thermotoga maritima*; *Nature* (*London*) **399** 323–329

Ochman H, Lawrence J G and Groisman E A 2000 Lateral gene transfer and the nature of bacterial innovation; *Nature* (*London*) **405** 299–304

Pan A, Dutta C and Das J 1998 Codon usage in highly expressed genes of *Haemophillus influenzae* and *Mycobacterium tuberculosis*: translational selection versus mutational bias; *Gene* **215** 405–413

Perez Luz S, Rodriguez-Valera F, Lan R and Reeves P R 1998 Variation of the ribosomal operon 16S-23S gene spacer region in representative *Salmonella enterica* subspecies; *J. Bacteriol.* **180** 2144–2151

Portnoy D A, Moseley S L and Falkow S 1981 Characterization of plasmids and plasmid-associated determinants of *Yersinia enterocolitica* pathogenesis; *Infect. Immun.* **31** 775–782

Reanney D C 1978 Coupled evolution: adaptive interactions among the genomes of plasmids, viruses, and cells; *Int. Rev. Cytol.* (*Suppl.*) **8** 1–68

Rivera M C, Jain R, Moore J E and Lake J A 1998 Genomic evidence for two functionally distinct gene classes; *Proc. Natl. Acad. Sci. USA* **95** 6239–6244

Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N and Yoshikawa M 1988 Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a; *J. Bacteriol.* **170** 2480–2484

Sassanfar M, Kranz J E, Gallant P, Schimmel P and Shiba K 1996 A eubacterial *Mycobacterium tuberculosis* tRNA synthetase is eukaryote-like and resistant to a eubacterial-specific antisynthetase drug; *Biochemistry* **35** 9996–10003

Schicklmaier P and Schmieger H 1995 Frequency of generalized transducing phages in Natural isolates of the *Salmonella typhimurium* complex; *Appl. Environ. Microbiol.* **61** 1637–1640

Sharp P M and Li W-H 1987 The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications; *Nucleic Acids Res.* **15** 1281–1295

Smith H O, Tomb I-F, Dougherty B A, Flerschmann R D and Venter J C 1985 Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome; *Science* **269** 538–540

Snel B, Bork P and Huynen M 1999 Genome phylogeny based on gene content; *Nat. Genet.* **21** 108–110

Sonea S and Paniset M 1976 Towards a new bacteriology; *Rev. Can. Biol.* **35** 103–167

Sueoka N 1988 Directional mutation pressure and neutral molecular evolution; *Proc. Natl. Acad. Sci. USA* **85** 2653–2657

Syvanen M and Kado C I 1998 *Horizontal gene transfer* (London: Chapman and Hall)

Whittam T S 1996 Genetic variation and evolutionary processes in natural populations *Escherichia coli*; *in Escherichia coli and Salmonella typhimurium: Cellular and molecular biology*, 2nd edition (eds) F C Neidhardt, R III Curtiss, J L Ingraham, E C C Lin, K B Low, B Magasanik, W S Reznikoff, M Riley, M Schaechter and H E Umbarger (Washington DC: American Society for Microbiology) pp 2708–2720

Woese C 1998 The universal ancestor; *Proc. Natl. Acad. Sci. USA* **95** 6854–6859

Wolf Y I, Aravind L and Koonin E V 1999a *Rickettsiae and Chlamydiae*: evidence of horizontal gene transfer and gene exchange; *Trends Genet.* **15** 173–175

Wolf Y I, Aravind L, Grishin N V and Koonin E V 1999b Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events; *Genome Res.* **9** 689–710

Yap W H, Zhang Z and Wang Y 1999 Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal gene transfer of an entire rRNA operon; *J. Bacteriol.* **181** 5201–5209