Lateral gene transfer and the nature of bacterial innovation

Howard Ochman*, **Jeffrey G. Lawrence**† **& Eduardo A. Groisman**‡

* *Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721-0088, USA*
† *Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
‡ *Howard Hughes Medical Institute, Washington University School of Medicine, Department of Molecular Microbiology, St Louis, Missouri 63110, USA*

....

Unlike eukaryotes, which evolve principally through the modification of existing genetic information, bacteria have obtained a significant proportion of their genetic diversity through the acquisition of sequences from distantly related organisms. Horizontal gene transfer produces extremely dynamic genomes in which substantial amounts of DNA are introduced into and deleted from the chromosome. These lateral transfers have effectively changed the ecological and pathogenic character of bacterial species.

Given that they are single-celled organisms and that their genome sizes vary by little more than an order of magnitude in length, bacteria display extraordinary variation in their metabolic properties, cellular structures and lifestyles. Even within relatively narrow taxonomic groups, such as the enteric bacteria, the phenotypic diversity among species is remarkable. Although the enteric bacteria as a whole share a myriad of traits denoting their common ancestry, each species also possesses a unique set of physiological characteristics that define its particular ecological niche (Fig. 1).

Several mechanisms could be responsible for the differences evident among bacterial species. Point mutations leading to the modification, inactivation or differential regulation of existing genes have certainly contributed to the diversification of microorganisms on an evolutionary timescale; however, it is difficult to account for the ability of bacteria to exploit new environments by the accumulation of point mutations alone. In fact, none of the phenotypic traits that are typically used to distinguish the enteric bacteria *Escherichia coli* from its pathogenic sister species *Salmonella enterica* can be attributed to the point mutational evolution of genes common to both[1]. Instead, there is growing evidence that lateral gene transfer has played an integral role in the evolution of bacterial genomes, and in the diversification and speciation of the enterics and other bacteria.

The significance of lateral gene transfer for bacterial evolution was not recognized until the 1950s, when multidrug resistance patterns emerged on a worldwide scale[2]. The facility with which certain bacteria developed resistance to the same spectrum of antibiotics indicated that these traits were being transferred among taxa, rather than being generated *de novo* by each lineage. Although the widespread impact of lateral gene transfer on bacterial evolution was not appreciated until much later, these early studies of rapid evolution by gene acquisition encompass four issues relevant to all current studies. First, how is it possible to detect and to identify cases of lateral gene transfer? Second, where do these genes come from, and by what mechanisms are they transferred? Third, what types of, and how many, traits have been introduced through lateral gene transfer? And fourth, at what relative rates are different classes of genes mobilized among genomes?

Detecting lateral gene transfer

How is it possible to establish whether a new trait, or a specific genetic region, is the result of horizontal processes? Naturally, it would be most satisfying to actually observe the conversion of a deficient strain in the presence of an appropriate donor—and all the more convincing to establish that genetic material had indeed been transferred and the manner in which it was acquired. But outside experimental settings—that is, for most cases of lateral gene transfer

in the recent or evolutionary history of a bacterial species—actual transfer events are only rarely observed, and unambiguous evidence of their occurrence must be derived from other sources.

Lateral gene transfer creates an unusually high degree of similarity between the donor and the recipient strains for the character in question. Furthermore, because each transfer event involves the introduction of DNA into a single lineage, the acquired trait will be limited to the descendents of the recipient strain and absent from closely related taxa, thereby producing a scattered phylogenetic distribution. However, lateral gene transfer need not be invoked to explain the sporadic occurrence of certain phenotypic traits, such as the ability to withstand particular antibiotics, because these properties can originate through point mutations in existing genes and therefore may evolve independently in divergent lineages. Thus, additional information is needed to discriminate between convergent evolution and lateral gene transfer. Clearly, the strongest evidence for (or against) lateral gene transfer derives from a molecular genetic analysis of the DNA sequences themselves.

DNA sequence information has been used in diverse ways to identify cases of lateral gene transfer, but the underlying basis of most applications is to discover features indicating that the evolutionary history of genes within a particular region differs from that of ancestral (vertically transmitted) genes. Similar to distinctive phenotypic properties, DNA segments gained through lateral gene transfer often display a restricted phylogenetic distribution among
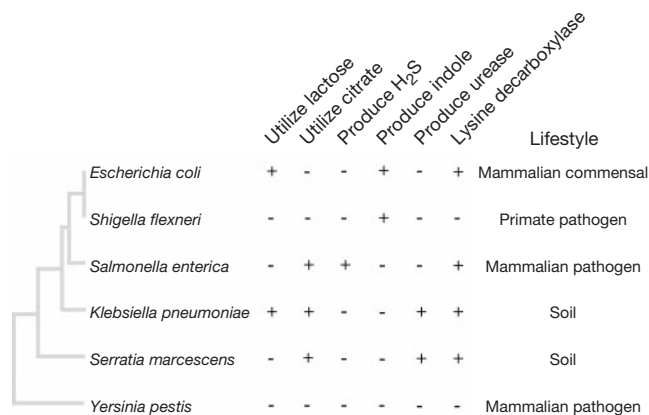
	Utilize lactose	Utilize citrate	Produce H$_2$S	Produce indole	Produce urease	Lysine decarboxylase	Lifestyle
Escherichia coli	+	−	−	+	−	+	Mammalian commensal
Shigella flexneri	−	−	−	+	−	−	Primate pathogen
Salmonella enterica	−	+	+	−	−	+	Mammalian pathogen
Klebsiella pneumoniae	+	+	−	−	+	+	Soil
Serratia marcescens	−	+	−	−	+	+	Soil
Yersinia pestis	−	−	−	−	−	−	Mammalian pathogen

Figure 1 Evolutionary relationships and phenotypic profiles of representative enteric bacteria. Plus denotes the presence and minus the absence of a trait in more than 85% of strains. Evolutionary relationships among species are based on nucleotide sequence information. In many cases, genes acquired by horizontal transfer confer the species-specific traits.

related strains or species. In addition, these species-specific regions may show unduly high levels of DNA or protein sequence similarity to genes from taxa inferred to be very divergent by other criteria[3]. The significance of aberrant phylogenies can be evaluated by phylogenetic congruency tests or other means[4].

Although gene comparisons and their phylogenetic distributions are useful for detecting lateral transfer, the DNA sequences of genes themselves provide the best clues to their origin and ancestry within a genome. Bacterial species display a wide degree of variation in their overall G+C content, but the genes in a particular species' genome are fairly similar with respect to their base compositions, patterns of codon usage and frequencies of di- and trinucleotides[5–7]. Consequently, sequences that are new to a bacterial genome, in other words, those introduced through horizontal transfer, retain the sequence characteristics of the donor genome and thus can be distinguished from ancestral DNA[1].

It is not surprising that genomic regions often manifest several attributes that denote their acquisition through lateral gene transfer. For example, a large number of *S. enterica* genes that are not present in *E. coli* (or any other enteric species) have base compositions that differ significantly from the overall 52% G+C content of the entire chromosome[8,9]. Within *S. enterica*, certain serovars (that is, lineages that exhibit a distinct composition of flagellar and/or lipopolysaccharide surface antigens) may contain more than a megabase of DNA not present in other serovars, as assessed by a genomic subtraction procedure. The base compositions of these anonymous serovar-specific sequences suggest that at least half were gained through horizontal transfer[10]. In addition to information obtained from the sequences of the genes themselves, the regions adjacent to genes identified as being horizontally transferred often contain vestiges of the sequences affecting their integration, such as remnants of translocatable elements, transfer origins of plasmids or known attachment sites of phage integrases, which further attest to their foreign origin in the genome.

Genetic exchange within and between bacterial species also acts upon homologous sequences, and numerous techniques have been developed to detect such events from sequence data[11]. However, the action of mismatch correction systems greatly reduces the efficiency of homologous recombination when donor and recipient sequences contain nonidentical bases[12,13]. As a result, homologous recombination is most successful in integrating DNA into the chromosome when the donor and recipient are relatively closely related. Because this type of genetic exchange principally affects the variation in existing genes, rather than introducing unique traits to the genome, its role in the ecological and physiological diversification of bacteria is apt to be negligible.

The scope of lateral gene transfer

The analysis of individual genes has uncovered numerous cases of lateral gene transfer; however, the ability to recognize horizontally acquired regions on the basis of their sequence characteristics makes it possible to assess the total proportion of foreign genes within a genome without resorting to gene phylogenies, sequence alignments or homology searches. Early attempts to establish the extent of laterally transferred sequences in a genome were limited to the very few microorganisms for which there was sufficient sequence information to get an unbiased sample of genes. Using this approach, it was originally estimated that between 10% and 16% of the *E. coli* chromosome arose through lateral gene transfer[9,14,15]— a range similar to the amount of unique DNA in the *E. coli* chromosome inferred from early alignments of the *E. coli* and *S. enterica* genetic maps[16].

The availability of complete genomic sequences provides an opportunity to measure and compare the cumulative amount of laterally transferred sequences in diverse bacterial genomes (Fig. 2). Potentially foreign genes are identified by their atypical nucleotide compositions, or patterns of codon usage bias[1,9]; after correcting for genes whose atypical features are due to amino-acid composition, the remaining genes are likely to have been introduced relatively recently by lateral gene transfer. Wide variation has been observed in the size and organization of 19 genomes analysed, and the amount of horizontally acquired DNA—represented as those open reading frames (ORFs) whose sequence characteristics depart from the prevalent features of their resident genome—ranges from virtually none in some organisms with small genome sizes, such as *Rickettsia prowazekii*, *Borrelia burgdorferi* and *Mycoplasma genitalium*, to nearly 17% in *Synechocystis* PCC6803. In all cases, very ancient horizontal transfer events, such as those disseminating transfer RNA synthetases[17], would not be detected using these methods.

In some species, a substantial proportion of horizontally transferred genes can be attributed to plasmid-, phage- or transposon-related sequences (Fig. 2, yellow bars). As observed in *E. coli*[1], a significant fraction of acquired DNA in *Synechocystis* PCC6803, *Helicobacter pylori* and *Archaeoglobus fulgidus* is physically associated with mobile DNA, which probably mediated the integration of these sequences into the chromosome. Although analyses based on sequence features indicate that large portions of bacterial genomes are attributable to lateral transfer, these methods can underestimate the actual number of transferred genes because sequences acquired from organisms of similar base compositions and codon usage patterns, as well as ancient transfer events, will escape detection.

Comparisons of completely sequenced genomes verify that bacteria have experienced significant amounts of lateral gene transfer,
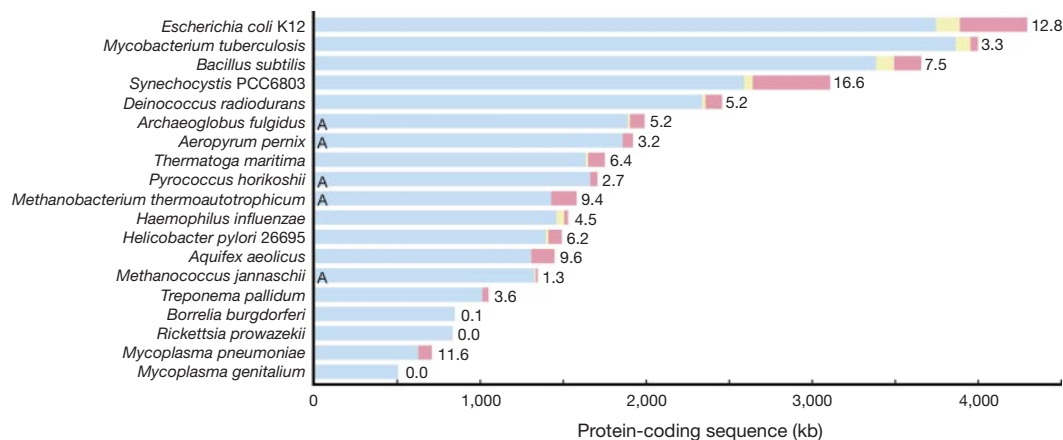

Figure 2 Distribution of horizontally acquired (foreign) DNA in sequenced bacterial genomes. Lengths of bars denote the amount of protein-coding DNA. For each bar, the native DNA is blue; foreign DNA identifiable as mobile elements, including transposons and bacteriophages, is yellow, and other foreign DNA is red. The percentage of foreign DNA is noted to the right of each bar. 'A' denotes an Archaeal genome.

resulting in chromosomes that are mosaics of ancestral and horizontally acquired sequences. The hyperthermophilic Eubacteria *Aquifex aeolicus* and *Thermotoga maritima* each contain a large number of genes that are most similar in their protein sequences and, in some cases, in their arrangements, to homologues in thermophilic Archaea. Twenty-four per cent of *Thermotoga*'s 1,877 ORFs and about 16% of *Aquifex*'s 1,512 ORFs, display their highest match (as detected by BLAST) to an Archaeal protein, whereas mesophiles, such as *E. coli*, *B. subtilis* and *Synechocystis*, have much lower proportions of genes that are most similar to Archaeal homologues[18,19].

Because the detection of lateral transfer by such systematic, gene-by-gene analyses depends upon the occurrence, phylogenetic positions and evolutionary rates of homologues in the currently available databases[20], the specific genes, as well as the overall amount of transferred DNA, recognized by this approach could differ from those resolved by examining atypical sequence characteristics. Moreover, the utility of BLAST similarity scores to infer gene ancestry depends upon the number of acknowledged homologues, the evolutionary relationships of organisms represented in the databases and the persistence of genes in a genome; hence, values derived from whole-genome comparisons will be refined as additional genomes are completed[21]. However, both methods reveal the potential for bacterial genomes to incorporate large numbers of unique sequences, often from very divergent organisms.

How and what sequences are acquired

In contrast to the evolution of new traits through the modification of existing sequences, the origin of new abilities through lateral gene transfer has three requirements. First, there needs to be a means for the donor DNA to be delivered into the recipient cell. Second, the acquired sequences must be incorporated into the recipient's genome (or become associated with an autonomous replicating element). And third, the incorporated genes must be expressed in a manner that befits the recipient microorganism. The first two steps are largely indiscriminate with respect to the specific genes or functional properties encoded by the transferred regions, and can occur through three mechanisms: transformation, transduction and conjugation.

(1) Transformation involves the uptake of naked DNA from the environment and has the potential to transmit DNA between very distantly related organisms. Certain bacterial species, such as *Neisseria gonorrhoeae* and *Haemophilus influenzae,* are perpetually competent to accept DNA, whereas others, such as *Bacillus subtilis* and *Streptococcus pneumoniae*, become competent upon reaching a certain physiological stage in their life cycle[22].

Effective transformation in *N. gonorrhoeae* and *H. influenzae* requires the presence of specific recognition sequences (5′-GCCGTCTGAA-3′ and 5′-AAGTGCGGT-3′, respectively), which are present in their respective genomes at frequencies far greater than those expected at random[23–25]. Although the presence of specific uptake sequences enhances the transformation efficiency between related species (for example, from *H. parainfluenzae* to *H. influenzae*), many of the naturally competent bacterial species, such as *B. subtilis* and *S. pneumoniae*, do not display sequence preference and are capable of high levels of transformation[22,26]. But transformation proficiency does not necessarily translate into correspondingly high rates of interspecific gene transfer: although species of *Haemophilus* are naturally competent, the *H. influenzae* Rd genome bears little foreign DNA beyond two large prophages (Fig. 2).

(2) New genetic material can be also introduced into a bacterium by a bacteriophage that has replicated within a donor microorganism and packaged random DNA fragments (generalized transduction) or the DNA adjacent to the phage attachment site (specialized transduction). The amount of DNA that can be transferred in a single event is limited by the size of the phage capsid, but can range upwards of 100 kilobases (kb). Although phages are prevalent in the environment[27,28], the spectrum of microorganisms that can be transduced depends upon receptors recognized by the bacteriophage. Like transformation, transduction does not require donor and recipient cells to be present at the same place, or even the same time. On the other hand, phage-encoded proteins not only mediate the delivery of double-stranded DNA into the recipient cytoplasm, but can also promote the integration of DNA into the chromosome and protect the transferred sequences from degradation by host restriction endonucleases.

(3) Conjugation involves physical contact between donor and recipient cells and can mediate the transfer of genetic material between domains (for example, between bacteria and plants, and between bacteria and yeast)[29,30]. Typically, DNA is transferred from a donor to a recipient strain by either a self-transmissible or mobilizable plasmid. Conjugation can also mediate the transfer of chromosomal sequences by plasmids that integrate into the chromosome (forming an Hfr strain that can transfer donor DNA to recipient cells), and by conjugative transposons, which encode proteins required for their excision from the donor, formation of a conjugative bridge and transposition into the recipient strain.

Despite the diversity of mechanisms mediating genetic exchange among prokaryotes, the introduction of DNA into a recipient cell's cytoplasm does not ensure successful gene transfer unless the transferred sequences are stably maintained in the recipient microorganism. DNA assimilation into the bacterial genome can exploit one of a number of processes including: (1) persistence as an episome, which requires selection to avoid stochastic loss; (2) homologous recombination, which will serve primarily to reassort variation among closely related taxa and is unlikely to allow introduction of novel traits; (3) integration mediated by bacteriophage integrases or mobile element transposases; and (4) illegitimate incorporation through chance double-strand break repair, as postulated for the integration of mitochondrial sequences into the yeast genome[31].

Through these mechanisms, virtually any sequence—even those originating in eukaryotes or Archaea—can be transferred to, and between, bacteria. The relatively small sizes of bacterial genomes imply, however, that either the rate of transfer or the maintenance of transferred sequences is very low, or that the maintenance of horizontally transferred sequences is offset by the loss of resident sequences. Bacterial genomes do not contain arbitrary assortments of acquired genes, and the plethora of documented transfer events has provided an abundance of information about the origins and functions of the acquired sequences. Naturally, these studies are biased towards identifying genes that impart both new and consequential functions upon the recipient organism; however, they illustrate both the unprecedented range of mechanisms by which traits are disseminated among bacteria and the impact of lateral gene transfer on bacterial evolution.

Traits introduced through lateral gene transfer

Antibiotic resistance. Antimicrobial resistance genes allow a microorganism to expand its ecological niche, allowing its proliferation in the presence of certain noxious compounds. From this standpoint, it is not surprising that antibiotic resistance genes are associated with highly mobile genetic elements, because the benefit to a microorganism derived from antibiotic resistance is transient, owing to the temporal and spatial heterogeneity of antibiotic-bearing environments. Plasmids are readily mobilizable between taxa and represent the most common method of acquiring antibiotic resistance determinants. As plasmids are rarely integrated into the chromosome, the acquired traits must confer an advantage sufficient to overcome not only inactivation by mutation, but also elimination by segregation.

Transposable elements can also promote the transfer of resistance genes between bacterial genomes. When a resistant determinant, or any gene conferring a selectable phenotype, is flanked by two

insertion sequences, it may be mobilized as a complex transposon; for example, two copies of the insertion sequence IS10 flank a tetracycline resistant determinant and regulatory gene to form transposon Tn10 (ref. 32). Likewise, two IS50 elements flank a three-gene operon that confers resistance to kanamycin, bleomycin and streptomycin forming transposon Tn5 (ref. 33), which can integrate into the chromosomes of phylogenetically diverse bacterial species.

Antibiotic resistance genes can also be propagated by integrons, which are gene expression elements that incorporate promoterless genes, thereby converting them into functional genes. Integrons consist of three elements (Fig. 3): an attachment site where the horizontally acquired sequence is integrated; a gene encoding a site-specific recombinase (that is, integrase); and a promoter that drives expression of the incorporated sequence[34,35]. A site-specific recombination event between the attachment site and a recombination site, known as the 59-base element, located downstream of a promoterless resistance gene results in the incorporation of the resistance gene and its expression from the integron promoter. Integrases exhibit specificity for their attachment sites but not for the 59-base elements, which allows the stockpiling of many resistance determinants within a given integron. Mobility of integrons demands capture by insertion sequences, transposons or conjugative plasmids[36]. Aside from antibiotic resistance genes, integrons have recently been implicated in the acquisition of virulence determinants by the cholera-causing bacterium, *Vibrio cholerae*[37].

Virulence attributes. Unlike the acquisition of antibiotic resistance, adoption of a pathogenic lifestyle usually involves a fundamental change in a microorganism's ecology. The sporadic phylogenetic distribution of pathogenic organisms has long suggested that bacterial virulence results from the presence (and perhaps acquisition) of genes that are absent from avirulent forms. Evidence for this view has taken several forms, ranging from the discovery of large 'virulence' plasmids in pathogenic *Shigella* and *Yersinia*[38–41] to the ability to confer pathogenic properties upon laboratory strains of *E. coli* by the experimental introduction of genes from other species[42,43].

Recent studies have discovered that horizontally acquired 'pathogenicity islands' are major contributors to the virulent nature of many pathogenic bacteria[44,45]. These chromosomally encoded regions typically contain large clusters of virulence genes and can, upon incorporation, transform a benign organism into a pathogen. Many pathogenicity islands are situated at tRNA and tRNA-like loci, which appear to be common sites for the integration of foreign sequences. For example, the tRNA^{selC} has repeatedly served as the integration site of pathogenicity islands in enteric bacteria[46], including the 70-kb PAI-1 of uropathogenic *E. coli*[47], the 35-kb LEE island of enteropathogenic *E. coli*[43], the 24-kb SHI-2 island of *Shigella flexneri*[48,49] and the 17-kb SPI-3 island of *S. enterica*[50] (Table 1).

The sequences flanking pathogenicity islands frequently contain short direct repeats reminiscent of those generated upon integration of mobile genetic elements, and ORFs within certain pathogenicity islands display sequence similarity to bacteriophage integrases. Several phages, including φR73 and P4 of *E. coli*, P22 of *S. enterica* and HP1 of *H. influenzae*, insert at or near tRNA genes, suggesting that pathogenicity islands are transferred and acquired through phage-mediated events, or that their incorporation involves the action of conserved integrases[51,52]. This is clearly the case in *Staphylococcus aureus*, in which a generalized transducing phage promotes the excision, replication and mobilization of a pathogenicity island harbouring the gene for toxic shock toxin[53].

Certain bacteriophages encode virulence determinants within their genomes, and lysogenization by such a bacteriophage results in the 'conversion' of a strain to a pathogenic variant. For example, the genes encoding exotoxin A in *Streptococcus pyogenes*[54], Shiga toxin in enterohaemorrhagic strains of *E. coli*[55], and the SopE GTP/GDP exchange factor necessary for host cell invasion by *S. enterica*[56] are carried by converting bacteriophages. In *V. cholerae,* the cholera toxin genes are encoded within the genome of a filamentous bacteriophage (termed CTXφ), which uses as its receptor the intestinal colonization factor TcpA encoded within the VP1 pathogenicity island[57]. VP1 constitutes the genome of another bacteriophage (which is distinct from CTXφ), however, and TcpA specifies the coat protein of this phage[58], such that acquisition of cholera toxin and, hence, virulence requires prior lysogenization by the VP1 phage.

Not only have pathogens originated through the acquisition of sequences by lateral gene transfer, but, in some cases, virulence also depends upon the absence of resident genes that diminish pathogenic potential (Fig. 4). An early example of such a virulence suppressor in enteric bacteria is the surface protease OmpT, which is present in nonpathogenic *E. coli* but absent from *Shigella*. The presence of OmpT attenuates virulence by interfering with the expression of the *Shigella* VirG protein, which is required for intercellular spread[59]. In addition to lacking *ompT*, Shigellae also have deletions of the region containing the *cadA* gene, which encodes lysine decarboxylase (Fig. 4). When the *cadA* gene from a benign strain of *E. coli* is introduced into *S. flexneri*, the resulting strain can no longer induce the fluid secretion normally associated with infection[60].
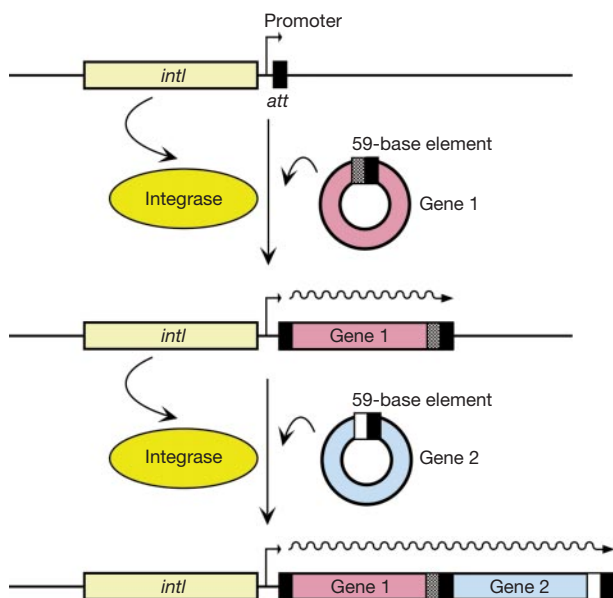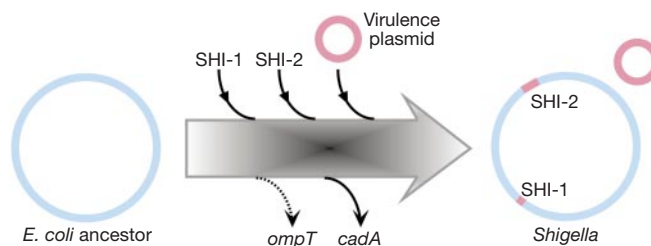

Figure 3 Gene capture and expression by integrons. Integrons consist of an attachment site (*att*), a gene encoding a site-specific recombinase (*intl*) and a promoter that drives expression of the incorporated sequence. Incorporation of a resistance gene (gene 1), and its expression from the integron promoter, result from a site-specific recombination event between the attachment site and a recombination site (known as the 59-base element) located downstream of a promoterless resistance gene. The integrase can recognize different 59-base elements, allowing the stockpiling of many resistance determinants within a given integron.

Figure 4 Succession of genetic events contributing to virulence in *Shigella*. Strains of *Shigella* are derived from *E. coli*, and pathogenicity in *Shigella* requires the acquisition of a virulence plasmid and two chromosomally encoded regions as well as the absence of two genes detected in *E. coli*. The deletion of *ompT* is displayed with a dashed line because it is not known whether all *E. coli* ancestors to *Shigella* harbour this phage-borne gene.

Metabolic properties. Lateral gene transfer has played a significant role in moulding bacterial genomes by mobilizing other physiological traits, which have, in effect, allowed recipient organisms to explore new environments. Although not surprising in retrospect, early alignments of the linkage maps of *E. coli* and *S. enterica* showed that many biochemical properties restricted to only one of these species, such as lactose fermentation by *E. coli* or citrate utilization by *S. enterica*, were encoded on chromosomal regions unique to each of these species[61]. Subsequent molecular genetic analyses have established that, in many cases, species-specific traits can be attributed either to the acquisition of genes through lateral gene transfer or to the loss of ancestral genes from one lineage. In this way, most ecological innovation in Eubacteria and Archaea is fundamentally different from diversification in multicellular eukaryotes.

The acquisition of new metabolic traits by horizontal transfer emphasizes the importance of natural selection as the arbiter of lateral gene exchange—the duration of acquired sequences will be fleeting if the genes do not contribute a useful or meaningful function upon introduction to the recipient cell. Therefore, we would expect that the successful mobilization of complex metabolic traits requires the physical clustering of genes, such that all necessary genes will be transferred in a single step. As a result, horizontal inheritance will select for gene clusters and for operons, which can be expressed in recipient cells by a host promoter at the site of insertion[62]. In this manner, *E. coli*, by acquiring the *lac* operon, gained the ability to use the milk sugar lactose as a carbon source and to explore a new niche, the mammalian colon, where it established a commensal relationship.

Whole-genome comparisons have uncovered sets of genes that are restricted to organisms that have independently adapted to a common lifestyle, such as Archaeal and Bacterial hyperthermophiles[18] or the intracellular pathogens *Rickettsia* and *Chlamydia*[63]. The distribution of such sequences in phylogenetically divergent but ecologically similar microorganisms has been ascribed to lateral gene transfer, and it is tempting to suggest that these genes, for which functions are presently unknown, contribute to the metabolic and physiological requirements of these unique environments.

Rate of sequence acquisition

The rate of DNA acquisition by the *E. coli* chromosome was measured indirectly by examining the amelioration of atypical sequence characteristics (for example, nucleotide composition) towards the equilibrium values displayed by this genome[1,9]. Such methods allow the estimation of the time of arrival for each segment of foreign DNA detected in the genome and, hence, the rate of successful horizontal genetic transfer. Taking into account the inevitable deletion of genes—bacterial genomes are not growing ever larger in size—horizontal transfer has been estimated to have introduced successfully ~16 kb per million years into the *E. coli* genome.

As evident from Fig. 2, there is broad variation in the amount of acquired DNA among bacterial genomes. Although some organisms may have only a limited capacity to transfer and exchange DNA, bacterial lifestyles can also contribute to the lower rates of gene acquisition in genomes containing small amounts of foreign DNA. For example, the intracellular habitat of *R. prowazekii* and *M. genitalium* probably shields the organism from exposure to potential gene donors and the opportunity to acquire foreign sequences.

The impact of acquired DNA

Lateral gene transfer provides a venue for bacterial diversification by the reassortment of existing capabilities. Yet, while the emergence of new phenotypic properties through lateral gene transfer furnishes several advantages, it also presents several problems to an organism. Newly acquired sequences, especially those conferring traits essential to only a portion of the bacterial life cycle, are most useful when they are appropriately and coordinately regulated with the rest of the genome. In *Salmonella*, the expression of several independently acquired virulence genes is under the control of a single regulatory system—the PhoP/PhoQ two-component system—that was already performing essential functions in the genome before the acquisition of these genes[50,64–66]. Although the precise manner by which each of these genes is regulated has yet to be resolved, these findings suggest that the physiological capabilities and adaptation of very divergent bacteria rely on a common set of universally distributed regulatory signals.

Despite wide diversity in the structure, organization and contents of bacterial genomes[67], there is relatively narrow variation in genome sizes. Evidence from experimental studies shows that bacterial genomes are prone towards deleting non-essential DNA, and the small, reduced genomes of host-dependent bacteria, such as *Mycoplasma*, *Chlamydia* and *Rickettsia*, attest to the tendency for

Table 1 Transfer RNA loci targeted by horizontally acquired DNA sequences

tRNA locus	Organism	Horizontally acquired DNA	Trait	Size (kb)
selC	*Escherichia coli*	Phage ΘR73	Phage genome	13
selC	Uropathogenic *E. coli*	PAI-1	Haemolysin	70
selC	Enteropathogenic *E. coli*	LEE	Type III secretion system; intimin receptor protein Tir	35
selC	*Salmonella enterica*	SPI-3	Macrophage survival protein MgtC; Mg^{2+} transporter MgtB	17
selC	*Shigella flexneri*	SHI-2	Aerobactin iron transport	24
leuX	Uropathogenic *E. coli*	PAI-2	Haemolysin; prf fimbriae	190
leuX	*E. coli*	Phage P4	Phage genome	
leu	*Haemophilus influenzae*	Phage HPI	Phage genome	
leu	*Bacteroides* sp.	NBU1		10
pheR	Uropathogenic *E. coli*	PAI-5	Haemolysin; prs fimbriae; cytotoxic necrotizing factor type 1	110
pheV	Uropathogenic *E. coli*	PAI-4	Haemolysin; pap fimbriae	170
pheV	*E. coli*	CTn*scr94*	Sucrose utilization	100
phe	*Mesorhizobium loti*	Symbiosis island	Nodulation and nitrogen fixation	500
asnT	*Yersinia enterocolitica*	HPI	Yersiniabactin	45
asnT	*Y. pseudotuberculosis*	HPI		
valV	*S. enterica*	SPI-2	Type III secretion system	40
serT	*S. enterica*	SPI-5	Inositol phosphate phosphatase	7
serV	*Dichelobacter nodosus*	Vap region	Putative toxin	12
metV	Uropathogenic *E. coli*		Haemolysin	50
ssrA	*D. nodosus*	Vrl region		27
ssrA	*Vibrio cholerae*	VP1	TcpA colonization factor and receptor for CTX phage	45
thr	*S. enterica*	Phage P22	Phage genome	44
thr	*Listeria ivanovii*	Inl region	Internalins C and D	5
arg	*Corynebacterium diphtheriae*	Corynephages	Phage genome; toxin	36
glyV	*Pseudomonas* sp.	*clc* element	Chlorocatechol-degradation	105

bacteria to delete the more expendable sequences from their genomes[68,69].

Because bacterial genomes can maintain only a finite amount of information against mutation and loss, chromosomal deletions will serve to eliminate genes that fail to provide a meaningful function, that is, the bulk of acquired DNA as well as superfluous ancestral sequences. Hence, bacterial genomes are sampling rather than accumulating sequences, counterbalancing gene acquisition with gene loss[70,71]. As a result, lateral gene transfer can redefine the ecological niche of a microorganism, which will, in effect, promote bacterial speciation.

A potential result of rampant interspecific recombination is the blurring of species boundaries, and the failure of any one gene to reflect the evolutionary history of the organism as a whole. Yet, as more bacterial genomes are examined, robust consensus phylogenies are being constructed from infrequently transferred genes, which provide the benchmark for gauging the scope and impact of lateral gene transfer. Rather than diminishing the utility of molecular phylogeny, the intermingling of genes and the resulting phylogenetic incongruities document the process of gene-transfer-mediated organismal diversification[3]. □

1. Lawrence, J. G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417 (1998).
2. Davies, J. Origins and evolution of antibiotic resistance. *Microbiologia* **12**, 9–16 (1996).
3. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129 (1999).
4. Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer Associates, Sunderland, Massachusetts, 1996).
5. Sueoka, N. On the genetic basis of variation and heterogeneity in base composition. *Proc. Natl Acad. Sci. USA* **48**, 582–592 (1962).
6. Muto, A. & Osawa, S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl Acad. Sci. USA* **84**, 166–169 (1987).
7. Karlin, S., Campbell, A. M. & Mrázek, J. Comparative DNA analysis across diverse genomes. *Annu. Rev. Genet.* **32**, 185–225 (1998).
8. Groisman, E. A., Saier, M. H. Jr & Ochman, H. Horizontal transfer of a phosphatase gene as evidence for the mosaic structure of the *Salmonella* genome. *EMBO J.* **11**, 1309–1316 (1992).
9. Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**, 383–397 (1997).
10. Lan, R. & Reeves, P. R. Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* **13**, 47–55 (1996).
11. Maynard Smith, J. The detection and measurement of recombination from sequence data. *Genetics* **153**, 1021–1027 (1999).
12. Vulic, M., Dionisio, F., Taddei, F. & Radman, M. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in Enterobacteria. *Proc. Natl Acad. Sci. USA* **94**, 9763–9767 (1997).
13. Rayssiguier, C., Thaler, D. S. & Radman, M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342**, 396–399 (1989).
14. Medigue, C., Rouxel, T., Vigier, P., Henaut, A. & Danchin, A. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.* **222**, 851–856 (1991).
15. Whittam, T. S. & Ake, S. in *Mechanisms of Molecular Evolution* (eds Takahata, N. & Clark, A. G.) 223–246 (Japan Scientific Society Press, Tokyo, 1992).
16. Riley, M. & Anilionis, A. Evolution of the bacterial genome. *Annu. Rev. Microbiol.* **32**, 519–560 (1978).
17. Wolf, Y. I., Aravind, L., Grishin, N. V. & Koonin, E. V. 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 (1999).
18. Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* **14**, 442–444 (1998).
19. Nelson, K. E. *et al.* Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 (1999).
20. Logsdon, J. M. Jr & Fuguy, D. M. *Thermotoga* heats up lateral gene transfer. *Curr. Biol.* **9**, R747–R751 (1999).
21. Gaasterland, T. Archaeal genomics. *Curr. Opin. Microbiol.* **2**, 542–547 (1999).
22. Dubnau, D. DNA uptake in bacteria. *Annu. Rev. Microbiol.* **53**, 217–244 (1999).
23. Goodman, S. D. & Scocca, J. J. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci. USA* **85**, 6982–6986 (1988).
24. Elkins, C., Thomas, C. E., Seifert, H. S. & Sparling, P. F. Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence. *J. Bacteriol.* **173**, 3911–3913 (1991).
25. Smith, H. O., Tomb, J. -F., Dougherty, B. A., Fleischmann, R. D. & Venter, J. C. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. *Science* **269**, 538–540 (1985).
26. Davison, J. Genetic exchange between bacteria in the environment. *Plasmid* **42**, 73–91 (1999).
27. Jiang, S. C. & Paul, J. H. Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* **64**, 2780–2787 (1998).
28. Schicklmaier, P. & Schmieger, H. Frequency of generalized transducing phages in natural isolates of the *Salmonella typhimurium* complex. *Appl. Environ. Microbiol.* **61**, 1637–1640 (1995).
29. Buchanan-Wollaston, V., Passiatore, J. E. & Canon, F. The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants. *Nature* **328**, 170–175 (1987).
30. Heinemann, J. A. & Sprague, G. F. J. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209 (1989).
31. Ricchetti, M., Fairhead, C. & Dujon, B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**, 96–100 (1999).
32. Kleckner, N. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 227–268 (American Society for Microbiology, Washington DC, 1989).
33. Berg, D. E. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 185–210 (American Society for Microbiology, Washington DC, 1989).
34. Hall, R. M. Mobile gene cassettes and integrons: moving antibiotic resistance genes in Gram-negative bacteria. *CIBA Found. Symp.* **207**, 192–205 (1997).
35. Rowe-Magnus, D. A. & Mazel, D. Resistance gene capture. *Curr. Opin. Microbiol.* **2**, 483–488 (1999).
36. Hall, R. M. & Collis, C. M. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.* **15**, 593–600 (1995).
37. Mazel, D., Dychinco, B., Webb, V. A. & Davies, J. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**, 605–608 (1998).
38. Portnoy, D. A., Moseley, S. L. & Falkow, S. Characterization of plasmids and plasmid-associated determinants of *Yersinia enterocolitica* pathogenesis. *Infect. Immun.* **31**, 775–782 (1981).
39. Maurelli, A. T., Baudry, B., d'Hauteville, H., Hale, T. L. & Sansonetti, P. J. Cloning of plasmid DNA sequences involved in invasion of HeLa cells by *Shigella flexneri*. *Infect. Immun.* **49**, 164–171 (1985).
40. Sasakawa, C. *et al.* Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a. *J. Bacteriol.* **170**, 2480–2484 (1988).
41. Gemski, P., Lazere, J. R., Casey, T. & Wohlhieter, J. A. Presence of virulence-associated plasmid in *Yersinia pseudotuberculosis*. *Infect. Immun.* **28**, 1044–1047 (1980).
42. Isberg, R. R. & Falkow, S. A single genetic locus encoded by *Yersinia pseudotuberculosis* permits invasion of cultured animal cells by *Escherichia coli* K-12. *Nature* **317**, 19–25 (1985).
43. McDaniel, T. K. & Kaper, J. B. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol. Microbiol.* **23**, 399–407 (1997).
44. Groisman, E. A. & Ochman, H. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**, 791–794 (1996).
45. Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**, 1089–1097 (1997).
46. Ritter, A. *et al.* tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. *Mol. Microbiol.* **17**, 109–121 (1995).
47. Blum, G *et al.* Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immun.* **62**, 606–614 (1994).
48. Moss, J. E., Cardozo, T. J., Zychlinsky, A. & Groisman, E. A. The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol. Microbiol.* **33**, 74–83 (1999).
49. Vokes, S. A., Reeves, S. A., Torres, A. G. & Payne, S. M. The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. *Mol. Microbiol.* **33**, 63–73 (1999).
50. Blanc-Potard, A. B. & Groisman, E. A. The *Salmonella selC* locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J.* **16**, 5376–5385 (1997).
51. Sun, J., Inouye, M. & Inouye, S. Association of a retroelement with a P4-like cryptic prophage (retronphage Theta;R73) integrated into the selenocystyl-tRNA gene of *Escherichia coli*. *J. Bacteriol.* **173**, 171–181 (1991).
52. Cheetham, B. F. & Katz, M. E. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.* **18**, 201–208 (1995).
53. Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29**, 527–543 (1998).
54. Weeks, C. R. & Ferretti, J. J. The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. *Infect. Immun.* **46**, 531–536 (1984).
55. Jackson, M. P., Neill, R. J., O'Brien, A. D., Holmes, R. K. & Newland, J. W. Nucleotide sequence analysis and comparison of the structural genes for Shiga-like toxin-I and Shiga-like toxin II encoded by bacteriophages from *Escherichia coli*. *FEMS Microbiol. Lett.* **44**, 109–114 (1987).
56. Mirold, S. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl Acad. Sci. USA* **96**, 9845–9850.
57. Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
58. Karaolis, D. K., Somara, S., Maneval, D. R. J., Johnson, J. A. & Kaper, J. B. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 675–679 (1999).
59. Nakata, N. *et al.* The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci. *Mol. Microbiol.* **9**, 459–468 (1993).
60. Maurelli, A. T., Fernández, R. E., Bloch, C. A., Rode, C. K. & Fasano, A. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 3943–3948 (1998).
61. Riley, M. & Sanderson, K. E. in *The Bacterial Chromosome* (eds Riley, M. & Drlica, K.) 85–96 (American Society for Microbiology, Washington DC, 1990).
62. Lawrence, J. G. & Roth, J. R. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**, 1843–1860 (1996).
63. Wolf, Y. I., Aravind, L. & Koonin, E. V. Rickettsiae and Chlamydia: evidence of horizontal gene transfer and gene exchange. *Trends Genet.* **14**, 442–444 (1998).
64. Groisman, E. A. & Ochman, H. How *Salmonella* became a pathogen. *Trends Microbiol.* **5**, 343–349 (1997).
65. Groisman, E. A. The ins and outs of virulence gene expression: Mg^{2+} as a regulatory signal. *Bioessays* **20**, 96–101 (1998).
66. Deiwick, J., Nikolaus, T., Erdogan, S. & Hensel, M. Environmental regulation of *Salmonella* pathogenicity island 2 gene expression. *Mol. Microbiol.* **31**, 1759–1773 (1999).
67. Casjens, S. The diverse and dynamic structure of bacterial genomes. *Annu. Rev. Genet.* **32**, 339–377 (1998).
68. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).
69. Andersson, J. O. & Andersson, S. G. E. Insights into the evolutionary process of genome degradation. *Curr. Opin. Genet. Dev.* **9**, 664–671 (1999).
70. Lawrence, J. G. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* **2**, 519–523 (1999).
71. Lawrence, J. G. & Roth, J. R. in *Organization of the Prokaryotic Genome* (ed. Charlebois, R. L.) 263–289 (American Society for Microbiology, Washington, 1999).