Lateral and oblique gene transfer

Howard Ochman

Sequence information from complete genomes, and from multiple loci of strains within species, is transforming the way that we investigate the evolution of bacteria. Such large-scale assessments of bacterial genomes have provided evidence of extensive gene transfer and exchange. Except in rare cases, these two processes do not seem to be coupled: certain species, such as *Escherichia coli*, undergo relatively low levels of gene exchange; but the emergence of pathogenic strains is associated with the acquisition of numerous virulence factors by lateral gene transfer.

Addresses
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721-0088, USA;
e-mail: hochman@email.arizona.edu

Current Opinion in Genetics & Development 2001, **11**:616–619

0959-437X/01/$ – see front matter

Abbreviations
MLEE multilocus enzyme electrophoresis
MLST multilocus sequence typing
MMR mismatch repair

Introduction

Recent literature on complete genomic sequences might leave the impression that gene transfer between organisms — even those assigned to different Kingdoms or Domains of life — is so pervasive that virtually any sequence can turn up in any lineage and the boundaries between species are breaking down. This is not the case. Whereas lateral transfer has certainly been a significant factor in the rapid adaptation and speciation of many bacterial lineages, the overall stability of the genome is, in fact, what allows one to assess the role of gene transfer.

Two types of genes are brought into the genome by lateral processes: those that are new to the organism and those that are homologous to existing genes. It is important to make this distinction because considering them as part of a single process has led to certain misconceptions about the impact of lateral gene transfer. Both types of transfer events influence the evolution of a lineage (as well as our notions about the relationships among organisms) but do so in a very different manner and context. In this review, I highlight several cases where lateral gene transfer has greatly influenced the biology, or on our interpretation of the biology, of an organism. Because actual transfer events are rarely observed, the evidence for horizontal gene transfer can be quite indirect. Therefore, also mentioned is a recent example of what might better be referred to as 'oblique' (i.e. not-quite-horizontal) gene transfer, where the support for gene exchange is less secure.

Ring out the old

The potential for acquiring (i.e. replacing) an existing gene generally decreases with the phylogenetic distance between the donor and recipient lineages [1,2]. Most gene transfer of this type occurs by homologous exchange, whose probability increases with genetic similarity. Aside from such mechanistic considerations, an acquired gene or gene segment is more likely to be beneficial — or at least not deleterious — if it is similar to the original sequence.

For more than two decades, one of the central issues investigated by bacterial population geneticists has concerned the role of recombinational exchange in the generation of allelic diversity within species [3]. Leaving aside concerns about how it might be possible to define bacterial species in light of varying degrees of genetic exchange, information about the amount of recombination, particularly within pathogenic species, has important ramifications about the manner by which bacteria might adapt to their hosts, respond to control, or be traced for epidemiological purposes [4,5].

Since the 1980s, the organization of genetic diversity within scores of bacterial species has been indexed by multilocus enzyme electrophoresis (MLEE), which discriminates among strains on the basis of the electrophoretically detectable allelic variation in numerous proteins [6]. But with the advent of rapid and economical sequencing came a shift towards the use of a multilocus sequence typing (MLST) scheme, which characterizes bacterial isolates according to the nucleotide sequences of roughly 450-bp fragments derived from several housekeeping genes (often corresponding to loci assayed originally by MLEE) [7].

Obtaining sequences of multiple loci usually assists in the identification and discrimination among strains; but by examining the congruence between phylogenetic trees reconstructed from individual genes, Feil *et al.* [8•] have assessed the extent of recombination in six species of pathogenic bacteria. Within four of the species — *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *Staphylococcus aureus* and *Neisseria menigitidis* — trees obtained for each of the genes did not match one another, indicative of recombinational exchange; however, in two species, pathogenic *E. coli* and *Haemophilus influenzae*, the recombination rates are low, as reflected by the congruence between the different gene trees.

Although *E. coli* populations are largely clonal, having experienced relatively low levels of gene exchange [6,9], recent focus has been directed towards the evolution of genes underlying the process methyl-directed mismatch repair (MMR) [10••,11]. The MMR system serves to block recombination in dissimilar sequences [12], and strains deficient in MMR display enhanced rates of mutations and gene

exchange, which are likely to be advantageous to bacteria that must survive heterogeneous environments or host responses to their presence [13–15].

When compared with phylogenies reconstructed from several other genes in natural isolates of *E. coli*, the trees based on the MMR genes present numerous incongruities that result from a relatively high level of gene exchange [10••,11]. Because high mutation rates might produce a long-term disadvantage by generating deleterious mutations, the high level of gene exchange apparent in MMR genes (owing, in part, to MMR mutants) might serve as a mechanism to restore function to defective MMR alleles, thereby relieving the consequences of increased mutation rates.

Tests of phylogenetic congruence can establish whether particular gene trees differ from one another, or from the phylogeny of organisms as a whole. Given the possibility of transferring virtually any sequence, however, it may be that certain genes are more representative of the actual evolutionary history of organisms and could thus be used as the backdrop onto which phylogenetic incongruities and events of horizontal transfer can be discovered. Most view ribosomal RNAs (rRNAs), and genes encoding conserved proteins, as being less susceptible to transfer because they are essential, universally distributed, and under stronger selective constraints resulting from their complex interactions with other proteins [16,17].

Despite recent claims to the contrary [18•], there are cases of phylogenetic ambiguity that have resulted from the transfer of rRNA [19–21]. Most bacteria accommodate many (up to 15) copies of rRNA genes, but their sequences are usually identical (or nearly so) within a genome, further supporting the reliability of these molecules as markers of organismal phylogeny. A few years ago, Yap *et al.* [21] reported the case of a thermophilic actinomycete (a high G+C Gram positive bacterium) that has two very distinct types of rRNA operons — a condition best explained by the acquisition of the alternative form from another actinomycete sharing a similar ecology.

A broad survey of the 16S and 23S rRNAs from additional actinomycetes shows that horizontal transfer might also be influencing the evolution of these genes on a different scale [21]. Not only are there cases in which the 16S and 23S genes yield different phylogenies, but it seems that lateral gene transfer has replaced small segments in the rRNAs of many actinomycete genera. This type of mosaicism, brought about by intragenic recombination and gene conversion, is usually restricted to genes within a bacterial species, and such events are also thought to contribute to the uniformity of rRNA operons within a genome.

Ring in the new

A second type of lateral transfer involves the acquisition of new genes, which, owing to their sporadic distribution, are generally not very useful for establishing phylogenetic affinities. More than in cases of homologous recombination and gene exchange, such transfer events often supply genes that confer novel phenotypic properties and have resulted in the rapid adaptation of some bacterial species.

Among the most notable examples of this form of evolution involve the conversion of benign and perfectly reasonable bacteria into pathogens after appropriating the proper virulence determinants [22,23]. The emergence of a pathogen can result from the acquisition of a single gene — not a rare event given the suspected amount of gene transfer among bacterial species — but in the case of many strains of pathogenic *E. coli*, virulence has resulted from an elaborate series of events that occurred independently in different lineages.

Shigella, the etiologic agent of bacillary dysentery, has long been known to be closely allied to *E. coli*; and, in fact, the Shigellae would more aptly be classified as pathogenic sublineages of *E. coli* because the genetic variation within the four species of *Shigella* is encompassed within the range detected in natural populations of *E. coli*. Furthermore, most *Shigella* species are polyphyletic; that is, each comprises isolates that arose from more than one ancestral lineage of *E. coli* [24–26]. The multiple independent origins of *Shigella* from different strains of *E. coli* are all the more surprising given the number and nature of genes that distinguish *Shigella* from nonpathogenic *E. coli*: all strains of *Shigella* carry a large virulence plasmid and at least two chromosomally encoded pathogenicity islands. And in addition to these acquired virulence elements, all Shigellae have deleted or inactivated genes whose products suppress virulence.

A similar situation has occurred in other enteropathogenic and enterohaemorragic strains of *E. coli*, including the virulent food-borne pathogen *E. coli* O157:H7. Reid *et al.* [27••] carried out an MLST study that shows that pathogenic *E. coli* strains are evolving more or less independently and not freely exchanging genes with one another. They super-imposed the incidence of several known virulence determinants onto the resulting strain phylogeny. The distribution provided clear evidence that virulence is a recent and derived condition, and that many existing *E. coli* lineages acquired the same virulence factors in parallel, suggesting that they have a selective advantage for accumulating certain combinations of factors that increase virulence.

The resonance of things past

The shear amount of data coming from genome sequencing projects, along with the still rather limited number of completely sequenced organisms, has led to the application of several alternative methods for inferring cases of lateral gene transfer. For example, recent elucidation of the complete sequence of *E. coli* O157:H7 yielded more than 1300 genes that were not present in the *E. coli* K-12 laboratory strain, and several of the larger O157-specific regions included phage-related sequences and were adjacent to transfer RNA loci that are known sites for the integration of

bacteriophage [28•]. Similarly, the genome sequence of *Streptococcus pyogenes* — the bacterial pathogen responsible for several human diseases, including toxic shock syndrome, scarlet fever and rheumatic fever — revealed several prophage-associated virulence regions [29].

Because lateral transfer can result in genes that display an unusually high degree of similarity to those in a distantly related lineage, the most common method of recognizing acquired genes has been through database searches. The closest matches from such searches are likely to be genes from other fully sequenced genomes because these constitute a disproportionate number of annotated but unassigned open reading frames, and because transfer events most probably involve novel, rather than universally distributed, sequences.

On the basis of BLAST searches, large proportions of the genomes of *Aquifex aeolicus* [30], *Thermotoga maritima* [31] and *Thermoplasma acidophilum* [32] comprise laterally transferred genes, with many of the acquired genes most similar in their protein sequences to those present in very distantly related organisms inhabiting similar environments. However, BLAST indices are based on regions of similarity between DNA or protein sequences [33], and the most closely related sequences arising from database searches need not be from the closest phylogenetic relative, partly due to the sporadic representation of sequenced organisms in current databases [34]. The same type of analysis of the complete human genome sequence yielded more than a hundred genes whose 'best-hit' similarities (i.e. BLAST scores) and absence from non-vertebrate eukaryotes seemed indicative of direct transfer from bacteria to vertebrates [35••]. Such events demand transfer into the vertebrate germ line to persist; therefore, although this is a relatively low number of acquired genes by bacterial standards, these findings are extraordinary. It should be noted, however, that these results are subject to numerous sources of error and the number of putative cases of bacteria to vertebrate transfer seems vastly overestimated [36••–38••].

Conclusions

Analysis of complete genomic sequences is bolstering the view that lateral gene transfer has had an extraordinary impact on bacterial genomes. But to what extent is this the consequence of the sample of microorganisms for which genome sequences are available and of the procedures used to infer lateral gene transfer? It is certain that some cases of gene transfer are tentative, especially among those identified solely on the basis of similarity searches, but the fact is that a bacterial genome completely unaffected by this mode of evolution and devoid of laterally transferred sequences would be rare. So far, the only genomes that seem to be 'frozen' are those of the human pathogen *Mycoplasma genitalium* and the aphid endosymbiont *Buchnera aphidicola*, both of which are very reduced in size and exist under ecological conditions that hinder gene acquisition and exchange [39–41]. However, the overwhelming majority of bacterial species do not reside in such environments, and have opportunities for contact with other microorganisms (either directly or through transmissible elements) and genomes that accommodate exogenous DNA.

Acknowledgements

The author acknowledges generous support from the National Institutes of Health and the Department of Energy.

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. 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* 1997, **94**:9763-9767.

2. Davison J: **Genetic exchange between bacteria in the environment.** *Plasmid* 1999, **42**:73-91.

3. Selander RK, Levin BR: **Genetic diversity and structure in *Escherichia coli* populations.** *Science* 1980, **210**:545-547.

4. Selander RK, Musser JM: **Population genetics and bacterial pathogenesis.** In *Molecular Basis of Bacterial Pathogenesis*, vol X. Edited by Iglewski BH, Clark VL. San Diego: Academic Press; 1990:11-36.

5. Ochman H: **The evolution of bacterial pathogens.** In *Principles of Bacterial Pathogenesis*. Edited by Groisman EA. New York: Academic Press, 2000:1-42.

6. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS: **Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics.** *Appl Environ Microbiol* 1986, **51**:873-884.

7. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA *et al.*: **Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms.** *Proc Natl Acad Sci USA* 1998, **95**:3140-3145.

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

As MLST gains broader use in investigating the population structure of bacteria, there will be an increasing need for studies such as these, which apply common analytical procedures to examine data from several species. These researchers use (and develop) state of the art methods, and are able to discover trends among diverse organisms.

9. Whittam TS: **Genetic variation and evolutionary processes in natural populations of *Escherichia coli*.** In Escherichia coli *and* Salmonella: *Cellular and Molecular Biology*, edn 2. Edited by Neidhardt FC *et al.* Washington: ASM Press; 1996:2708-2720.

10. •• Denamur E, Lecointre G, Darlu P, Tenaillon O, Acquaviva C, Sayada C, Sunjevaric I, Rothstein R, Elion J, Taddei F *et al.*: **Evolutionary implications of the frequent horizontal transfer of mismatch repair genes.** *Cell* 2000, **103**:711-721.

The extensive application of phylogenetic methods provides convincing evidence that certain portions of the genome undergo an alternate mode of evolution. This is among the first papers appearing in *Cell* that relies substantially upon phylogenetic reconstructions; could this be the beginning of a trend?

11. Brown EW, LeClerc JE, Li B, Payne WL, Cebula TA: **Phylogenetic evidence for horizontal transfer of *mutS* alleles among naturally occurring *Escherichia coli* strains.** *J Bacteriol* 2001, **183**:1631-1644.

12. Rayssiguier C, Thaler DS, Radman M: **The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch mutants.** *Nature* 1989, **342**:396-401.

13. Sniegowski PD, Gerrish PJ, Lenski RE: **Evolution of high mutation rates in experimental populations of *E. coli*.** *Nature* 1997, **387**:703-705.

14. Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B: **Role of mutator alleles in adaptive evolution.** *Nature* 1997, **387**:700-702.

15. LeClerc JE, Cebula TA: ***Pseudomonas* survival strategies in cystic fibrosis.** *Science* 2000, **289**:391-392.

16. Jain R, Rivera MC, Lake LA: **Horizontal transfer among genomes: the complexity hypothesis.** *Proc Natl Acad Sci USA* 1999, **96**:3801-3806.

17. Lawrence JG: **Gene transfer, speciation and the evolution of bacterial genomes.** *Curr Opin Microbiol* 1999, **2**:529-523.

18. • Kurland CG: **Something for everyone. Horizontal gene transfer in evolution.** *EMBO Rep* 2000, **1**:92-95.
Great entertainment. This is a feisty and opinionated attempt to dispel the notion of rampant lateral gene transfer among organisms.

19. Sneath PH: **Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences.** *Int J Syst Bacteriol* 1993, **43**:626-629.

20. Wang Y, Zhang Z: **Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes.** *Microbiology* 2000, **146**:2845-2854.

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

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

23. Groisman EA, Ochman H: **How to become a pathogen.** *Trends Microbiol* 1994, **2**:289-294.

24. Ochman H, Whittam TS, Caugant DA, Selander RK: **Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*.** *J Gen Microbiol* 1983, **129**:2715-2726.

25. Rolland K, Lambert-Zechovsky N, Picard B, Denamur E: ***Shigella* and enteroinvasive *Escherichia coli* strains are derived from distinct ancestral strains of *E. coli*.** *Microbiology* 1998, **144**:2667-2672.

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

27. •• Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS: **Parallel evolution of virulence in pathogenic *Escherichia coli*.** *Nature* 2000, **406**:64-67.
See annotation [28•].

28. • Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA *et al.*: **Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.** *Nature* 2001, **409**:529-533.
These two papers [27••,28•] consider, from the viewpoint of *E. coli*, what it takes to be pathogen. Over 25% of the genes in the O157 genome are not present in the previously sequenced non-pathogenic laboratory strain of *E. coli*, and a large number of the O157-specific genes are related to genes in other enteric pathogens. The wide use of *E. coli* as an experimental organism will allow the functions of many of these genes to be tested, and the emergence of pathogenic strains can be studied in the context of population genetics and evolutionary biology.

29. Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, Primeaux C, Sezate S, Suvorov AN, Kenton S: **Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.** *Proc Natl Acad Sci USA* 2001, **98**:4658-4663.

30. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV: **Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles.** *Trends Genet* 1998, **14**:442-444.

31. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC *et al.*: **Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*.** *Nature* 1999, **399**:323-329.

32. Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W: **The genome sequence of the thermoacidophilic scavenger, *Thermoplasma acidophilum*.** *Nature* 2000, **407**:508-513.

33. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**:3389-3402.

34. Koski LB, Golding GB: **The closest BLAST hit is often not the nearest neighbor.** *J Mol Evol* 2001, **52**:540-542.

35. •• International human genome sequencing consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
If it were possible to assign three stars, this paper, along with the one appearing simultaneously from the Celera group [42] would be the ones. Aside from describing the technical *tour de force*, both papers are beautifully written and organized, and manage to convey how scientists think about genomes. Required reading.

36. •• Salzberg SL, White O, Peterson J, Eisen JA: **Microbial genes in the human genome: lateral transfer or gene loss?** *Science* 2001, **292**:1903-1906.
See annotation [38••].

37. •• Andersson JO, Doolittle WF, Nesbo CL: **Are there bugs in our genome?** *Science* 2001, **292**:1848-1850.
See annotation [38••].

38. •• Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR: **Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates.** *Nature* 2001, **411**:940-944.
These three papers [36••–38••] discuss sequences in the human genome that were apparently introduced by lateral gene transfer from bacteria. Problems with gene identification and annotation, cases of probable gene loss, uneven rates of sequence evolution, the paucity of sampled genomes for comparisons, and the lack of phylogenetic signal all contribute to an overestimate in the true number of transferred genes.

39. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM *et al.*: **The minimal gene complement of *Mycoplasma genitalium*.** *Science* 1995, **270**:397-403.

40. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS.** *Nature* 2000, **407**:81-86.

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

42. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.*: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.