

# Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis

Howard Ochman\* and Nancy A. Moran

Traditionally, evolutionary biologists have viewed mutations within individual genes as the major source of phenotypic variation leading to adaptation through natural selection, and ultimately generating diversity among species. Although such processes must contribute to the initial development of gene functions and their subsequent fine-tuning, changes in genome repertoire, occurring through gene acquisition and deletion, are the major events underlying the emergence and evolution of bacterial pathogens and symbionts. Furthermore, pathogens and symbionts depend on similar mechanisms for interacting with hosts and show parallel trends in genome evolution.

Bacteria form intimate, and quite often mutually beneficial, associations with a variety of multicellular organisms. The diversity of these associations, combined with their agricultural and clinical importance, have made them a prominent focus of research. Of microbial genomes completed or under way, more than two-thirds are organisms that are either pathogens of humans or dependent on a close interaction with a eukaryotic host. But current databases and scientific literature present a distorted view of bacterial diversity. Estimates of bacterial diversity from various environmental sources, including the biota from animal surfaces and digestive tracts, show that pathogens represent a very small portion of microbial species (1–5). Potential hosts, especially humans with their broad geographic distribution and high population densities, are constantly besieged by bacteria in the environment, but most do not cause infections.

Not only are pathogenesis and symbiosis relatively rare among bacterial species, it is a derived condition within bacteria as a whole, as evident from the fact that bacteria existed well before their eukaryotic hosts. The appearance of the major groups of eukaryotes, whose diversification could proceed only after the origin of mitochondria by endosymbiosis (6), marks the initial availability of abundant suitable hosts. The mitochondria themselves derive from a single lineage within the alpha subdivision of the Proteobacteria (7); that is to say, they are nested near the tips of the overall bacterial phylogeny. Thus, the distribution of pathogens and symbionts in numerous divergent clades of bacteria reflects the repeated and independent acquisition of this life-style.

What then does it take to become a patho-

gen or symbiont? The basic requirements involve overcoming the numerous physical, cellular, and molecular barriers presented by the host. Typically, this entails contacting and entering the host body, growth and replication using nutrients from host tissues, avoidance of host defenses, persistence and replication, and finally exiting and infecting new hosts. Selection favors bacteria that achieve each of these steps, regardless of whether the ultimate outcome of the interaction is harmful, benign, or beneficial to the host. Only from the host's perspective are these distinctions crucial.

Because the biological processes needed to successfully infect hosts are largely the same for pathogens and symbionts, we can begin to ask whether these different bacteria have deployed common molecular mechanisms to initiate and maintain their relationships with hosts. Combining in-depth genetic and functional studies with data from comparative genomics is helping to address these issues by revealing unexpected gene homologies among organisms. Furthermore, these homologies can be used as a basis for inferring the functional roles of genes in the large number of pathogenic and symbiotic microbes that are not subject to experimental manipulation. Results from this approach implicate two ongoing and seemingly contradictory processes in bacterial genomes—gene acquisition and gene loss—as playing major roles in promoting the spectrum of interactions between bacteria and their hosts.

## Acquiring Genes Necessary for Host Interactions

Among the most successful methods for identifying the molecular genetic basis of virulence traits has been the comparison of pathogens to related nonpathogenic strains or species. These experiments have taken two general forms: in the first, genes from the pathogen are assayed for their ability to confer a virulence phenotype upon a normally avirulent strain, and in the second, genes

from pathogens are tested via mutational analysis for their role in virulence. Analysis of sequences recovered by these methods has made it evident that many of the genes required for virulence are restricted to pathogenic organisms and have been introduced into genomes by lateral transfer. Although point mutations may sometimes modulate a virulence phenotype (8), gene acquisition is much more prevalent as the basis for virulence evolution within lineages. This process is so pervasive that species-specific chromosomal regions containing virulence genes are now classed under the general heading of “pathogenicity islands” (9–11).

Pathogenicity islands can encompass very large genetic regions, sometimes spanning more than 100 kilobases (kb), and their frequent integration at or near tRNA loci suggests that many were introduced into bacterial genomes via phage-mediated transfer events (9, 11, 12, 13). This mobility would naturally give rise to a situation whereby homologous mechanisms for host interactions are adopted by very different microbes. Recent findings show that diverse plant and animal pathogens employ a broadly conserved set of genes, which encode a type III secretion apparatus used to deliver specialized proteins into host cells (14–16). Among the best illustrations that symbionts have similar evolutionary problems, as well as solutions, has come from the discovery of a 500-kb “symbiosis island” inserted at a tRNA locus in the plant symbiont *Mesorhizobium loti* (17). Perhaps even more surprising is the recent demonstration that an acquired type III secretion system is necessary to establish the symbiosis between *Sodalis glossinidius* and its tsetse fly host (18).

Despite its widespread role in bacterial diversification, the capacity for lateral gene transfer to convert an organism into a successful pathogen or symbiont is not indiscriminate. For *Escherichia coli*, a normal resident of the mammalian intestinal flora, the acquisition of a single pathogenicity island is sufficient to confer pathogenic properties upon benign strains (9, 11, 19). In *Salmonella enterica*, a closely related enteric, five separate islands are necessary for disease progression (20), and recent elucidation of the complete genomic sequence of enterohaemorrhagic *E. coli* O157:H7 revealed a multitude of strain-specific regions encoding virulence determinants (21). In these cases, the virulence phenotype demands horizontally ac-

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85718, USA.

\*To whom correspondence should be addressed. E-mail: hochman@email.arizona.edu

quired genes, implying that ancestral strains were not pathogenic before the acquisition of these genes. However, it is important to note that the ancestors to these pathogens already possessed many of the traits required to cope with environments presented by animal hosts, including mechanisms to counteract host defenses and biosynthetic pathways to compensate for nutrient deficiencies in the host. Thus, these organisms were preadapted to become pathogens upon the acquisition of additional virulence determinants encoded within pathogenicity islands. Based on mutational analyses of gene function, about 5% of the *Salmonella* genome has a role in virulence, and many of these genes are ancestral and are present in nonpathogenic strains of *E. coli* (20, 22). Therefore, the potential for virulence to arise through gene acquisition is mostly limited to bacterial lineages that have already been in long contact with eukaryotes.

Complete genome sequences offer new opportunities to evaluate gene content and the impact of lateral gene transfer on the evolution of pathogenesis and symbiosis. In particular, such analyses have extended our abilities to evaluate the role of gene transfer in the many organisms that cannot be cultured and studied experimentally. Most cases of lateral transfer uncovered from whole-genome comparisons involve sequences whose homologs are uniquely shared by phylogenetically divergent, fully sequenced organisms (23, 24). By this approach, the intracellular pathogens *Rickettsia prowazekii* and *Chlamydia trachomatis* were found to have exchanged genes thought to be needed for host interactions (25), suggesting that a shared niche promotes the transfer and maintenance of genes required for adapting to a particular host or life-style. Additional support for gene exchange occurring within a common host environment derives from evidence of the interspecies transfer of a virulence factor (*sodC*) between *Haemophilus influenzae* and *Neisseria meningitidis*, two pathogens that colonize the human respiratory tract (26).

There exist mechanisms in bacteria for transferring virtually any sequence between pairs of organisms spanning any degree of genetic relatedness (13, 27). However, the successful acquisition of genes requires not only the delivery and incorporation of DNA, but also the maintenance of acquired DNA within the recipient lineage. This maintenance depends on natural selection favoring the trait conferred by the acquired gene and preventing the spread of mutations that destroy its function. Thus, although DNA is continuously being transferred among lineages, the persistence of a gene within a genome is dependent on its utility to the recipient organism (28). Bacteria living in proximity to a host are more likely to benefit from the acquisition of

pathogenicity- or symbiosis-related genes.

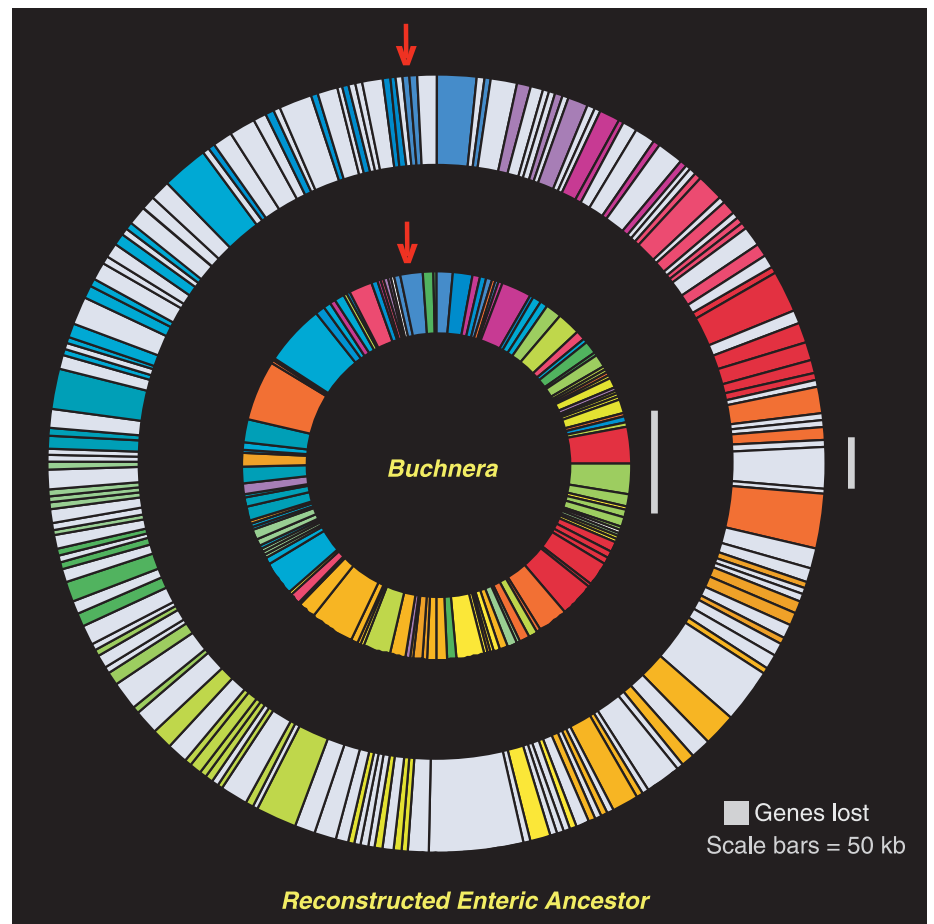
The addition of genes, through duplications and horizontal transfer, may also contribute indirectly to the emergence of opportunistic pathogens that are versatile with respect to environments and hosts. These organisms, exemplified by *Pseudomonas aeruginosa*, maintain large genomes containing an arsenal of genetic mechanisms for dealing with diverse environments, antimicrobial agents, and substrates, among them certain eukaryotic tissues (29).

**Gene Loss and Genome Degradation**

It is understandable that organisms with the capacity to acquire genes by lateral transfer would exploit this mechanism to evolve new traits, but it is not as obvious that deletion of genes could serve as a means of

bacterial adaptation. Strains of *Shigella* (the etiologic agents of dysentery) lack *ompT*, a gene present in closely related nonpathogenic *E. coli*. Introduction of *ompT*, which encodes a surface protease, suppresses *Shigella* virulence by disrupting intercellular spread (30); thus, in *Shigella*, the deletion of DNA was crucial to the development of virulent strains.

A large set of symbionts and pathogens have undergone more massive gene loss (Fig. 1). All of the smallest genomes for cellular organisms (in the range of about 0.6 to 1.5 Mb) belong to obligate pathogens or symbionts (31), and these small-genome organisms constitute a substantial proportion of eukaryote associates. Phylogenetic analyses indicate that these microbes are derived from ancestors with larger genomes (32–34) and that they belong to large



**Fig. 1.** Genome reduction in *Buchnera*-APS, the bacterial endosymbiont of aphids. *Buchnera* has undergone massive genome reduction (to 0.64 Mb) (34). Virtually all of its 590 genes have close homologs in the genomes of the enteric bacteria, which, along with phylogenetic evidence, indicates that its genome was derived from a much larger genome resembling modern enteric bacteria such as *E. coli*. In this depiction, the outer ring represents the hypothetical ancestral genome for enteric bacteria, produced by removing horizontally acquired regions [identified on the basis of phylogenetic distribution among enteric bacteria (51)] from the genome of *E. coli* MG1655. Gray bands denote ancestral sections of the genome that have been eliminated during the evolution of the *Buchnera* lineage. Colored bands represent regions within which ancestral gene arrangements persist in *Buchnera*, although many individual genes within these regions have been lost. *Buchnera* retains 21% of ancestral genes. The *Buchnera* genome is drawn at a larger scale than the ancestor (scale bars beside each genome are 50 kb), and the bands are colored to match corresponding ancestral regions. For each genome, the arrow indicates origin of replication.

and ancient clades consisting of only pathogens or symbionts (e.g., the Mollicutes, the Rickettsiae, the spirochetes, and the Chlamydiae). Many of these small-genome bacteria are intracellular, and all are distinguished by being able to replicate only in close association with a eukaryotic host.

Despite a consistent correlation between genome size and the obligate association with host cells, genome reduction is not simply an adaptive response to living within hosts. Instead, the trend toward large-scale gene loss reflects a lack of effective selection for maintaining genes in these specialized microbes (35). Because the host presents a constant environment rich in metabolic intermediates, some genes are rendered useless by adoption of a strictly symbiotic or pathogenic life-style. These superfluous sequences are eliminated through mutational bias favoring deletions, a process apparently universal in bacterial lineages (36). Thus, all of the fully sequenced small genomes display a pattern of loss of biosynthetic pathways, such as those for amino acids that can be obtained from the host cytoplasm (37–39).

Genome reduction also results partly from the loss of apparently beneficial genes: many of the eliminated genes encode proteins that enhance efficiency of universal cellular processes, including DNA repair, translation, and transcription (37–43). This loss, along with elevated rates of fixation of deleterious mutations within genes, arises from the inefficiency of natural selection due to the partitioning of populations among hosts and large fluctuations in population sizes (44). Chance fixation of such mutations destroys the function of beneficial (but not essential) loci that are consistently present in bacteria with larger genomes. The resulting pseudogenes shrink through successive DNA deletions and persist in degraded states for long periods of time, as exemplified in *Rickettsia* (36). Finally, small-genome symbionts and pathogens may often lack the opportunity or ability to incorporate foreign DNA, such that genome shrinkage is not countered by gene acquisition.

In summary, one of the most distinctive features of many symbiotic and pathogenic genomes, extremely small size, does not appear to be an adaptation for living within hosts but a neutral or even deleterious consequence of long-term evolution under the conditions imposed by these life-styles. One consequence of genome reduction is that specialized symbionts and pathogens are unable to reacquire the multitude of eliminated genes and thus cannot revert to a life-style independent from hosts. This irreversibility is supported by the phylogenetic distribution of small-genome organisms, which occur in clades that are uni-

formly symbiotic or parasitic (32, 33, 39).

Although pathogens and symbionts show clear parallels in their genetic responses to living within hosts, they differ in some aspects of their genome contents. Specialized mutualistic symbionts, often in cooperation with their hosts, are able to circumvent host defenses through mechanisms such as sequestration within specialized host tissues or cells that function as refuges (45). Indeed, genes encoding surface molecules potentially involved in host defenses are virtually absent in the genome of the mutualistic aphid endosymbiont *Buchnera* (34). In contrast, pathogens, even those with reduced genomes (e.g., *Rickettsia*, *Mycoplasma*, *Borrelia*, *Treponema*, and *Chlamydia*), possess substantial numbers of genes thought to function in cellular interactions and antigenic mechanisms (37, 38, 41–43).

Unlike pathogens, symbionts may devote part of their genomes to processes that are more directly beneficial to the host rather than to the bacterial cell itself. *Buchnera* retains and even amplifies genes for the biosynthesis of amino acids required by hosts, devoting almost 10% of its genome to these pathways, which are missing from pathogens with similarly small genomes (34). Because of their fastidious growth requirements, the biological role of obligately associated symbionts can rarely be determined experimentally (46, 47). However, genome comparisons can provide a means for determining their functions in hosts. Such future research should reveal, for example, whether the endosymbionts of blood-feeding hosts, such as *Wigglesworthia glossinia* in tsetse flies, retain pathways for biosynthesis of vitamins absent from blood (48), whether the symbiont *Vibrio fischeri* provides functions other than bioluminescence to its squid host (47), and whether the mutualistic *Wolbachia* of filarial nematodes (49) contain genes for host benefit that are absent in the parasitic *Wolbachia* of arthropods (50).

The evolutionary transitions underlying pathogenic and symbiotic life-styles are varied, but most entail gene transfer and gene loss occurring within bacterial lineages. These changes in genome content show that adaptation and diversification in bacteria occur by a process very different from that envisioned in classic models of evolution, which are based on the accumulation of small changes within individual genes. Given this mode of evolution in bacteria, full genome sequences are providing, for the first time, a means of illuminating the origins of diverse bacterial associates of eukaryotes.

References

1. E. F. Delong, *Trends Biotechnol.* **15**, 203 (1997).
2. P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **180**, 4765 (1998).

3. V. Torsvik, D. L. Daee, R. A. Sandaa, L. Overas, *J. Biotechnol.* **64**, 53 (1998).
4. M. R. Ranon, R. M. Goodman, J. Handelsman, *Trends Biotechnol.* **17**, 403 (1999).
5. I. Kroes, P. W. Lepp, D. A. Relman, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14547 (1999).
6. B. F. Lang, M. W. Gray, G. Burger, *Annu. Rev. Genet.* **33**, 351 (1999).
7. M. W. Gray, G. Burger, B. F. Lang, *Science* **283**, 1476 (1999).
8. E. V. Sokurenko et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8922 (1998).
9. E. A. Groisman, H. Ochman, *Cell* **87**, 791 (1996).
10. J. Hacker, G. Blum-Oehler, I. Muhldorfer, H. Tschape, *Mol. Microbiol.* **23**, 1089 (1997).
11. J. Hacker, J. B. Kaper, *Annu. Rev. Microbiol.* **54**, 641 (2000).
12. B. F. Cheetham, M. E. Katz, *Mol. Microbiol.* **18**, 201 (1995).
13. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
14. C. J. Hueck, *Microbiol. Mol. Biol. Rev.* **62**, 379 (1998).
15. L. W. Cheng, O. Schneewind, *Trends Microbiol.* **8**, 214 (2000).
16. J. E. Galán, A. Collmer, *Science* **284**, 1322 (2000).
17. J. T. Sullivan, C. W. Ranson, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5145 (1998).
18. C. Dale, S. A. Young, D. T. Haydon, S. C. Welburn, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1883 (2001).
19. T. K. McDaniel, J. B. Kaper, *Mol. Microbiol.* **23**, 399 (1997).
20. E. A. Groisman, H. Ochman, *Trends Microbiol.* **5**, 343 (1997).
21. N. T. Perna et al., *Nature* **409**, 529 (2001).
22. F. Bowe et al., *Infect. Immun.* **66**, 3372 (1998).
23. L. Aravind, R. L. Tatusov, Y. I. Wolf, D. R. Walker, E. V. Koonin, *Trends Genet.* **14**, 442 (1998).
24. K. E. Nelson et al., *Nature* **399**, 323 (1999).
25. Y. I. Wolf, L. Aravind, E. V. Koonin, *Trends Genet.* **15**, 173 (1999).
26. J. S. Kroll, K. E. Wilks, J. L. Farrant, P. R. Langford, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12381 (1998).
27. J. Davison, *Plasmid* **42**, 73 (1999).
28. J. G. Lawrence, *Curr. Opin. Microbiol.* **2**, 529 (1999).
29. C. K. Stover et al., *Nature* **406**, 959 (2000).
30. N. Nakata et al., *Mol. Microbiol.* **9**, 459 (1993).
31. S. Casjens, *Annu. Rev. Genet.* **32**, 339 (1998).
32. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
33. J. Maniloff, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10004 (1996).
34. S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, *Nature* **407**, 81 (2000).
35. S. G. E. Andersson, C. G. Kurland, *Trends Microbiol.* **6**, 263 (1998).
36. J. O. Andersson, S. G. E. Andersson, *Curr. Opin. Genet. Dev.* **9**, 664 (1999).
37. C. M. Fraser et al., *Science* **270**, 397 (1995).
38. S. G. Andersson et al., *Nature* **396**, 133 (1998).
39. N. A. Moran, J. J. Wernegreen, *Trends Ecol. Evol.* **15**, 321 (2000).
40. S. Razin, S. D. Yogev, Y. Naot, *Microbiol. Mol. Biol. Rev.* **62**, 1094 (1997).
41. C. M. Fraser et al., *Science* **281**, 375 (1998).
42. C. M. Fraser et al., *Nature* **390**, 580 (1997).
43. R. S. Stephens et al., *Science* **282**, 754 (1998).
44. N. A. Moran, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2873 (1996).
45. P. Buchner, *Endosymbiosis of Animals with Plant Microorganisms* (Wiley, New York, 1965).
46. R. W. Lee, J. J. Robinson, C. M. Cavanaugh, *J. Exp. Biol.* **202**, 289 (1999).
47. K. L. Visick, J. Foster, J. Dojno, M. McFall-Ngai, E. G. Ruby, *J. Bacteriol.* **182**, 4578 (2000).
48. X. Chen, S. Li, S. Aksoy, *J. Mol. Evol.* **48**, 49 (1999).
49. C. Bandi, B. Slatko, S. L. O'Neill, *Parasitol. Today* **15**, 428 (1999); M. J. Taylor, H. F. Cross, K. Biló, *J. Exp. Med.* **191**, 1429 (2000).
50. R. Stouthamer, J. A. Breeuwer, G. D. Hurst, *Annu. Rev. Microbiol.* **53**, 71 (1999).
51. M. McClelland et al., *Nucleic Acids Res.* **28**, 4974 (2000).