Rates and patterns of chromosome evolution in enteric bacteria
Howard Ochman* and Ulfar Bergthorsson†

Although several types of large-scale alterations potentially affect the structure and organization of bacterial genomes, recent analyses of physical maps and complete genomic sequences reveal that chromosome heterogeneity in enteric bacteria has resulted from the acquisition and deletion of large segments of DNA. These acquired sequences can provide novel functions immediately upon their introduction and play a significant role in the diversification of bacterial species.

Addresses
*Department of Biology, University of Rochester, Rochester, NY 14627, USA; e-mail: hoch@uhura.cc.biology.rochester.edu
†Department of Biology, University of Utah, Salt Lake City, UT 84112, USA
Correspondence: H Ochman

Current Opinion in Microbiology 1998, 1:580–583
http://biomednet.com/elecref/1369527400100580
© Current Biology Ltd ISSN 1369-5274

Abbreviation
sv. serovar

Introduction
How quickly, and in what way, do bacterial chromosomes evolve? Comparisons of the complete genomic sequences of some dozen microorganisms have revealed that certain features of bacterial chromosomes have remained conserved over vast evolutionary periods [1*,2*]. For example, there are similarities in the arrangement of sequences, and in the order of specific genes surrounding the replication origin in diverse bacterial genomes; in fact, in lieu of direct experimental localization, the same organization of these specific genes in other organisms has allowed the preliminary positioning of replication origins [3,4**]. In addition, the proteins encoded in each of the completely sequenced genomes can be classified according to their structure, function and evolutionary relationships, yielding information about the ancestry of groups of proteins, the origins of novel sequences, and the modification of gene function [5**,6].

Although the comparative analysis of whole chromosomes is extremely useful for identifying the unique and conserved features of genomes, the organisms sequenced to date are too distantly related, and their genomes too divergent, to allow for a reasonable assessment of the rates and patterns of chromosome evolution in bacteria. This is not surprising: organisms occupying very different habitats and separated for hundreds of millions of years have sustained so many changes that the evolutionary histories may have been erased [2*,7*,8*]. Even the congeneric species Mycoplasma pneumoniae and M. genitalium (for which complete genome sequences are available) are sufficiently different in chromosome size and organization to impede characterization of the specific order, number, and rate of genetic events that distinguish their genomes [9]. Therefore, to trace the course of chromosome evolution in bacteria, it is most informative to compare the genomes of very closely related organisms. In this paper, we review several recent studies have detected unprecedented amounts of chromosome variation within bacterial species and have begun to identify the factors that influence genome size and organization.

Comparative genomics of Escherichia coli and Salmonella
Early alignments of the genetic maps of Escherichia coli K12 and Salmonella enterica serovar (sv.) Typhimurium LT2 revealed a surprising amount of congruence and promoted the view that bacterial chromosomes are evolutionarily well conserved. But although the chromosomes of these species are of similar size and gene order, there are several large regions unique to each of the species as well as differences in their gene arrangements. These findings have led to investigations focusing on four general aspects of chromosome evolution: first, the degree of chromosome heterogeneity within species; second, the mechanisms generating diversity in chromosome organization; third, the role of variable regions; and, finally, the rate of chromosome evolution.

Despite the similarity in chromosome size of laboratory isolates — the E. coli K12 chromosome is 4.6 Mb and the Typhimurium LT2 chromosome is 4.8 Mb — strains from natural populations of each of these species can differ by as much as 1 Mb (i.e. more than 20% in total length). Chromosome sizes in natural isolates of E. coli range from 4.5 to 5.5 Mb, whereas the chromosomes from strains of S. enterica range from 3.9 to 4.9 Mb in length [10,11*,12*]. Though of equivalent size ranges, the apportionment of chromosome diversity appears to differ in these two species. Nearly all of the chromosome size variation in Salmonella was detected among strains typed to a single serovar, S. enterica sv. Typhi, and much of the size diversity in serovars other than Typhi maps to a single region of the chromosome. In contrast, the chromosome size variation in E. coli is not restricted to a single group of strains, or to a specific region of the chromosome, and there is a large phylogenetic component to the variation, as evident from the significant differences in chromosome size between broad subspecific groups [13].

Source of chromosome diversity
Under laboratory conditions, enteric bacteria display a high incidence of spontaneous gene inversions, duplications and translocations; however, these events do not seem to account for much of the chromosome heterogeneity observed in natural populations. There are a few
cases where strains are known to differ by chromosome rearrangements: for example, many of the differences in the organization of *S. enterica* sv. Typhi chromosome have resulted from recombination between ribosomal RNA operons [14], and *E. coli* K12 and *S. enterica* sv. Enteriditis SSU7798 can each be distinguished from *Typhimurium* LT2 on the basis of a large inversion spanning the replication terminus [15••]. Natural strains of *E. coli* and *S. enterica* undoubtedly harbor additional rearrangements that are not detected by large-scale physical mapping techniques; however, there is overwhelming evidence that the majority of variation in genome size and content is generated through the acquisition and deletion of large chromosomal segments.

Base composition is relatively homogeneous over the entire bacterial chromosome and, therefore, regions acquired through horizontal transfer from distantly related organisms can often be identified by their atypical base compositions or other sequence characteristics. Applying a genomic subtraction procedure, Lan and Reeves [16] estimated the amount of DNA contained in *S. enterica* sv. *Typhimurium* LT2 that was not present in the genomes of four divergent strains of *S. enterica*. They established that the *Typhimurium* LT2 genome contained as much as 1.3 Mb of unique DNA when compared to a very distantly related strain; by analyzing the genetic content and base composition of these strain-specific sequences, they concluded that nearly half of this DNA was gained through horizontal transfer.

Considerable attention has been directed towards the investigation of pathogenicity islands, which are chromosomal regions required for virulence in pathogenic strains but absent from related strains that do not cause disease [17,18]. Boyd and Hartl [19••] found that genes associated with pathogenicity islands were confined to those subgroups of *E. coli* with larger genome sizes, indicating that these regions contribute to the chromosome size variation in natural populations of *E. coli*. Most notably, the phylogenetic distribution of these virulence-associated genes within *E. coli* suggests that pathogenicity islands are ancestral to these subgroups, despite the fact that most of these present-day strains were originally isolated from healthy hosts.

Chromosomal regions recognized as arising through horizontal transfer are commonly situated adjacent to tRNA loci [20]. This distribution implicates bacteriophages as vehicles for gene transfer, since several lysogenic coliphages target tRNA loci, presumably because tRNA sequences are conserved across taxa. The same tRNA locus has been used as the integration site for distinct DNA segments in different strains or species. For example, the *selC* locus contains a 35 kb insert in enteropathogenic *E. coli* [21•], a 70 kb insert in uropathogenic *E. coli* [18], and a 17 kb insert in *S. enterica* [22•], and each of these inserted regions encodes unique sequences and was acquired independently. Because of the variety of sequences already detected at tRNA loci, it is certain that further analysis of these sites will uncover new inserts in other strains or species.

**Functional aspects of chromosome heterogeneity**

Many changes in chromosome structure and organization have deleterious effects on DNA replication and cell growth. In laboratory populations of *E. coli*, there is a reduction in cell fitness related to the degree of asymmetry in the location of the replication terminus relative to the origin [23], and in *Salmonella*, large inversions encompassing certain regions of the chromosome are lethal to the cell [24]. Although selection on the basis of chromosome structure has certainly occurred in natural populations — for example, each of the inversions distinguishing *E. coli* sv. Enteritidis and sv. Typhimurium encompasses the terminus but maintains chromosome symmetry — there is currently no evidence that any of the observed differences in chromosome size and organization have an effect on the fitness of strains recovered from natural populations. Since bacterial chromosomes contain mostly coding regions, it has been predicted that strains with smaller genomes would lack scores of genes and would be adapted to rapid growth under complex nutrient-rich conditions, whereas strains with larger genomes would be at a selective advantage in nutrient-poor conditions [12•]. But among natural isolates of *E. coli*, there is no association between growth rate and genome size, and other factors, such as the translational efficiency of ribosomes, outweigh the effects of total genome size on growth rates [25].

Because much of the heterogeneity in *E. coli* and *S. enterica* chromosomes is attributed to insertions and deletions, the evolutionary consequences of this variation are best judged from the specific functions of genes residing in these regions. Although most of the acquired or deleted regions probably have little effect on the organism, some horizontally acquired sequences can provide a novel function immediately upon their introduction and, in effect, change the character of a species. The genetic contents of several species-specific regions are known, and many confer novel metabolic properties that distinguish *E. coli* and *S. enterica*. For example, the *lac* operon, allowing for the degradation of β-galactosides, was acquired by *E. coli*, and in *S. enterica* the *cob* and *pdu* operons, providing for vitamin B12 biosynthesis and the B12-dependent degradation of propanediol, also arose through horizontal transfer [26••].

As noted, many of the genes implicated in pathogenesis — such as the genes required for host cell invasion and intramacrophage survival by *Salmonella* [17,21••,27,28] — occur on unique segments of the chromosome. In contrast to pathogenicity islands, which can confer virulence upon their acquisition, the absence of certain chromosomal regions may also contribute to pathogenesis. For example, *Shigella* and
enteroinvasive *E. coli* each harbor a chromosomal deletion of up to 90 kb relative to *E. coli* K12. This region inhibits enterotoxin activity and its deletion facilitates the expression of plasmid-borne virulence genes in these pathogens [29].

### Rates of chromosome evolution

Despite the relatively high frequency of spontaneous alterations in bacterial chromosomes, evidence from natural isolates of *E. coli* and *S. enterica* suggests that large-scale changes in the size and organization of bacterial genomes are relatively rare events on an evolutionary timescale. Short-term rates of chromosome evolution have been analyzed in experimental populations of *E. coli* B propagated for 2000 generations under four different thermal regimes: 32°C, 37°C, 42°C, and alternating 32°C and 42°C. In an analysis of these lines, chromosome alterations were monitored by pulsed-field gel electrophoresis and a total of 12 changes (formed through deletions, duplications, inversions or point mutations) were detected, with none occurring in strains cultured at 37°C (U Bergthorsson, unpublished data). These results suggest that strains propagated at temperatures other than 37°C either have higher mutation rates or have incurred changes that are adaptive under non-standard growth conditions.

How does this rate of chromosomal evolution in bacteria compare to that in eukaryotes? Differences in genome size and organization make such comparisons difficult; but for experimental populations of *Saccharomyces cerevisiae*, the number of chromosome alterations, as detected by pulsed-field gel electrophoresis [30], was about twofold higher than in *E. coli*.

### Conclusions

Despite their longer evolutionary history, bacteria display far less variation in total genome size than do eukaryotes, but they are certainly more adept at obtaining new chromosomal regions that confer unique phenotypic properties. Based on the sequence characteristics of horizontally transferred genes, it has been estimated that *E. coli* has acquired new sequences at a rate of nearly 30 kb per million years since diverging from the *Salmonella* lineage 100 million years ago. Although many of the acquired genes were subsequently deleted, those sequences that have persisted, which constitute about 18% of the current genome, include genes that distinguish *E. coli* from other enteric species [31**]. Such findings promote the view that bacterial speciation is driven by horizontal transfer, which introduces genes permitting the rapid exploitation of new environments. The identification of the full complement of sequences that distinguish closely-related bacteria will undoubtedly come from comparisons of complete genome sequences. Such studies have, in fact, already begun with the recent completion of the nucleotide sequence of enterohemorrhagic *E. coli* 0157:H7, a pathogen whose chromosome is some 30% larger than that of *E. coli* K12.

### 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


This paper, along with [29], compares some of the completely sequenced genomes and describes the manner in which gene order and chromosome organization have evolved in bacteria. By first classifying genes into one of several functional classes and then examining their relative positions, the authors find that in both *Escherichia coli* and *Haemophilus influenzae* functionally related genes tend to be neighbors more often than functionally unrelated genes.


By examining the locations of homologous genes in several sequenced genomes, these authors detect very few evolutionarily conserved gene clusters and conclude that bacterial genomes are subject to repeated events that alter the chromosome structure.


Most bacterial chromosomes are polarized around their origins of replication, both in terms of base composition and gene distribution. This paper examines the extent of GC skew – the excess of G over C in the leading strand of replication – in several sequenced genomes and discusses the potential forces generating the strand asymmetry in base composition.


This is a superb analysis of the organization and evolution of bacterial chromosomes as reconstructed from complete genome sequences. The authors describe the relative rates at which differ features of the genome, ranging from protein sequence to overall chromosome content and structure, have changed over an evolutionary timescale.


A comparison of the complete sequences of several eubacterial and archael genomes shows that while families of proteins are conserved among taxa, gene order and gene families are not. On the basis of the heterogeneity among genomes, the authors conclude that horizontal transfer plays a significant role in the evolution of genomes.


On the basis of an analysis of the genes common to *E. coli* and *H. influenzae*, these authors hypothesize that the common ancestor of these species had a genome size similar to that of present day *E. coli*. Several gene duplications are ancestral and common to both species; except for few small regions, gene order has not been conserved since these species diverged.


On the basis of multilocus enzyme electrophoresis data, natural isolates of Typhi are thought to be genetically homogeneous; however, this study shows that strains of Typhi exhibit extraordinary plasticity in chromosome size and organization, with genomes that vary from 3.9 to 4.9 Mb in length.

Among natural isolates of *E. coli*, chromosome sizes range from 4.5 to 5.5 Mb. The distribution of this length variation is not random: strains with larger chromosomes are primarily found in certain subspecific groups and the chromosome size variation has a symmetric distribution with respect to the replication origin.


The distribution of four virulence determinants associated with pathogenicity islands is clustered in the subspecific groups of *E. coli* having larger genomes. The distribution suggests that these virulence genes were present in the ancestors of these subgroups despite the fact that the majority of these strains were originally isolated from healthy hosts.


The insertion of the LEE island can, in a single step, convert a benign strain of *E. coli* into a pathogen. This paper provides the newly determined sequence of the entire insert; and aside from containing genes that are apparently specific to LEE, this region also encodes a type III secretion apparatus homologous to those present in pathogenicity islands of other enteric pathogens, including *Salmonella*, *Shigella* and *Yersinia*.


The salC locus has been shown to be a common target for pathogenicity islands in *Escherichia coli*, and this study takes the clever approach of surveying the corresponding site in the *Salmonella* genome. The authors discover a 17 kb insert required for virulence, adding to the growing list of pathogenicity islands in *Salmonella*.


This paper very convincingly makes the case that bacterial genes specifying a single metabolic function are typically arranged into operons because this organization often facilitates efficient horizontal transfer of genes among organisms. Their model accounts for the mosaic structure of some bacterial genomes whereby ancestral chromosomal material is interspersed with horizontally transferred operons providing novel metabolic functions.


Strains of *Shigella* and enteroinvasive *E. coli* were found to have chromosomal deletions of cadA and the surrounding region. The presence of cadA inhibits enterotoxin activity and attenuates virulence in *Shigella* and these deletions are likely to have played a large role in the evolution of these pathogens. A similar situation exists for ompT, a surface protease which is present in *E. coli* K12 but not in *Shigella* and enteroinvasive *E. coli*.


Based on their GC contents, genes acquired through horizontal transfer can be distinguished from ancestral DNA. At the time of introduction, horizontally transferred genes reflect the base composition of the donor genome; but, over time, these sequences will ameliorate to reflect the DNA composition of the new genome. This paper develops a method to estimate the time of acquisition of horizontally transferred genes, which makes it possible to establish the age of transferred genes and the rate at which DNA has been acquired on an evolutionary timescale.