The Bacterial Chromosome
Edited by N. Patrick Higgins

Chapter 1

Where's the Beef? Looking for Information in Bacterial Chromosomes

JOHN R. ROTH

This chapter is a collection of thoughts, questions, and problems to consider while reading this book. It is not intended as a comprehensive review or a unifying theory for chromosome structure. In looking at the structure or sequence of a genome, one sees all aspects of its information content, regardless of how they got there or how they are maintained. Many of these features are difficult to understand as functions contributing to the fitness of a single cell. Some aspects of chromosome structure that are shaped by selfish behavior with few conventionally maintained features that may be understandable only at a population level are pointed out. The principles described here are well known but are often forgotten because we are so accustomed to identifying functions genetically (mutations with phenotypes) or biochemically (proteins with interesting properties). Genetics is good at revealing genes and other sequence elements that affect the phenotype of a single cell and that are likely to be maintained by conventional selection. Genome sequences reveal all information, even that maintained by aspects of population biology that are unlikely to help us understand chromosome behavior. Some conserved features may be positively noxious. The challenge to those trying to explain chromosome behavior is to identify the critical sequence elements (i.e., "the beef").

THE QUESTIONS

1. How do bacteria compact their chromosomes?

2. How do bacterial chromosomes move? Do chromosomal sites, perhaps recognizable as sequence elements, serve as attachment points for cytoskeletal motor proteins? Where in the cell are these fibers attached?

3. How is gene expression and its regulation integrated with chromosome replication, compaction, and movement?

APPROACHING THE QUESTIONS

An advantage of chromosome work in bacteria is the large number of complete genome sequences that are available for comparison. This body of data should help identify aspects of sequence that are universal, overrepresented, and perhaps critical for compaction and movement of chromosomes. However, in extracting this information, we must learn to distinguish the critical features from features that are unlikely to be mechanistically informative but may be interesting in their own right.

FORCES THAT DICTATE THE SIZE OF CHROMOSOMES

One might expect that a genome would be under constant selective pressure to expand, adding new genes that broaden the capabilities of the organism. Three aspects of population biology and lifestyle limit this expansion: mutation rate, population size, and recombination frequency. These principles were suggested by Muller (15, 35, 48, 49). First, if all other factors remain constant, mutation rate limits genome size; beyond a critical point, the genome can expand only if the mutation rate drops. This is because more information can be kept under selection if fewer mutations occur. (With no mutations, a huge genome

John R. Roth • Microbiology Section, Division of Biological Sciences, Center for Genetics and Development, University of California, Davis, Davis, CA 95616.

could be maintained without selection.) Second, larger populations allow genome expansion (with other factors constant) because selection is more effective (drift is less significant) and genes with a smaller fitness contribution can be maintained. Third, sexual recombination allows genome expansion by permitting assembly of intact information sets from those damaged by mutation. Expressed differently, selection works better on individual genes if recombination continuously rearranges allele combinations. Together these considerations suggest that (if other factors are held constant) genome size will be inversely related to mutation rate and directly related to population size and recombination frequency (35). Each organism or population may satisfy these relationships in different ways, but the values remain interdependent. (Organisms with a very high mutation rate generally have small genomes and large population sizes.) While the exact mathematical function is unknown, these general rules of Muller reflect the problems of maintaining genes in the face of continued mutation pressure. However, these forces may work differently on different genes.

A basic assumption made here is that each gene is individually subject to loss by mutation and can counteract that loss by individual strategies. That is, different genes compete with each other for maintenance in a chromosome (35). A gene remains in the chromosome (and avoids loss by mutation and drift) conventionally if its function enhances fitness; mutations that impair such a function reduce fitness and are removed from the population by selection. Some genes enhance their ability to compete with other genes by private (selfish) strategies that do not enhance fitness and may actually be detrimental to the organism.

INTRODUCTION TO SELFISH BEHAVIOR

Much of sequence analysis is based on the expectation that conservation implies a selectively valuable function that can be unraveled and explained. Most chromosome features do contribute to cell fitness and provide a conventionally selectable phenotype; these are likely to be reliable clues to mechanism. However, other sequence features may provide no increase in fitness to the organism or may actually be deleterious, but are conserved by private (selfish) selection schemes.

Selfish features can be "addictive"—they do little or nothing when present, but cause a fitness loss when removed by mutation. Other selfish features are "infectious"; that is, they spread (by replication or horizontal transfer) faster than they can be removed from the genome (and population) by counterselection or simple mutation and drift. It is suggested below that high copy number may be an additional distinct form of selfish behavior. Distinguishing selfish from conventional sequence elements can be difficult because elements can be maintained by a combination of conventional and selfish selective forces. Some conventional genes may enhance their ability to stay in the chromosome by minor use of a selfish mechanism. Some primarily selfish elements may enhance their survival by modestly enhancing cell fitness. In extreme cases, elements may be maintained by selfish mechanisms alone.

A point to keep in mind is that selfish behavior improves a gene's ability to compete with others for a position in the genome, but it does not make that gene more beneficial. Genes maintained by selfish behavior still impose a cost on aggregate fitness of the organism. Selfish behavior allows a gene to remain despite its inadequate or even negative contributions to fitness. Selfish elements take a free ride, but ultimately their fare has to be paid by conventional genes—you could not have a genome made up entirely of selfish elements.

CONSERVATION OF CHROMOSOME GENE ORDER: GENE POSITION MAY BE ADDICTIVE

One might expect that a particular gene would work equally well regardless of its chromosomal position. This idea is supported by the observation that chromosome map order is not well conserved over extensive evolutionary distances (50), suggesting that order is not under strong selection. However, order seems to be conserved on a shorter time scale. *Salmonella* and *Escherichia coli* have maintained essentially the same map since their divergence about 140 million years ago. Since rearrangements are mechanistically possible in these organisms, this conservation suggests that gene order might be under some kind of selection. However, gene order could also be an addictive feature of chromosome structure.

Genes can, in principle, work anywhere, but once their promoters are tuned for any particular location, their transplantation causes a loss of fitness. We presume that the strength of each promoter is selectively adjusted to optimize gene expression and that this optimization necessarily takes into consideration the local influences on expression. One local characteristic may be the degree of supercoiling that is characteristic of the site (see below); another may be the gene dosage status imposed by the position a gene occupies vis-à-vis the origin of replication.

When chromosome replication requires a significant fraction of the cell cycle, as is frequently the case for bacteria, average dosage of a particular gene varies with its position in the chromosome (distance from the origin). Genes replicated early are present in multiple doses for much of the cell cycle, while those replicated late have two copies only for the period between their replication and cell division. Dichotomous replication during rapid growth (when new forks are initiated before completion of the previous round) can increase this dosage gradient. Optimization of promoters and regulatory elements will necessarily take into consideration the time-averaged copy number of each gene.

Expression of a gene could be optimized for any location in the chromosome, but once that optimization is achieved, moving the gene to a new site would render suboptimal the tuning of its promoter. In essence, gene order might be an addictive property of chromosome structure restraining rearrangement. Even if tuning of an individual gene is only modestly affected by an inversion or translocation, the accumulated small fitness costs incurred by the many genes involved in a rearrangement could allow selection to efficiently eliminate individuals with an altered gene order.

Evidence that these changes occur is that expression of a particular gene changes as it is moved farther from the replication origin (61). Furthermore, evolution appears to have exploited this fact because genes with high expression levels (judged by a high codon adaptation index) tend to be located near the replication origin (64, 65).

Conservation of gene order is often attributed to lack of mechanisms for rearrangements or to the need to maintain the proper spacing of sequence features that are critical to the packing or movement of chromosomes. These could be factors, but the above arguments suggest that conservation could equally well be attributed to the "addictive" effects of well-tuned promoters.

OPERONS AND OTHER GENE CLUSTERS

A notable feature of bacterial chromosomes is the clustering of genes that contribute to a single selectable phenotype (11). Some of these clusters are operons (30), groups of genes cotranscribed from a single promoter, while others are made up of closely linked but independently expressed genes. Initially the evolution of operons was ascribed to economies of coregulation (29) or to setting the molecular ratio of proteins that acted together (1). More recently, this clustering has been ascribed to a selfish property of gene clusters (36) that gives them an advantage (over identical unclustered versions) in the larger context of horizontal transfer between genomes of different bacteria. This model suggests that clustering gives genes a selfish property of the infectious sort. Previously, Wheelis and Stanier suggested that horizontal transfer might drive clustering of different operons controlling parts of the same pathway (74).

The earliest regulatory explanations of operon evolution were called into question because the benefits proposed to explain operon formation would accrue only after genes were successfully brought together. It was hard to visualize how two genes could be brought together by slow steps with progressive increases in selective value, as expected for most evolutionary processes. Furthermore, the energy or material saved by coregulating genes in an operon seemed very small, once it was realized that a regulatory binding site of only a few bases is sufficient to allow effective coregulation of distant genes. Finally, the regulatory models did not explain clustering of independently transcribed genes with related functions.

The selfish operon model suggests that clustering provides no selective benefit to the cell (36). Instead the clustered state is a selfish property of genes that contributes to the ease with which they can spread horizontally between conspecific individuals or species. If clustering improves the likelihood that a set of genes moves horizontally, it may provide an advantage (to the genes themselves, not to their hosts) vis-à-vis identical genes in an unclustered state. The advantage conferred on genes by clustering accrues without any required improvement in the fitness of the organism (outlined below in more detail). If this is true, then operons and gene clusters may be examples of a conserved chromosome feature that does not tell us much about the physiological functions within an individual bacterial cell. The origin and maintenance of this feature is rather explained by horizontal gene transfer and population biology. Occasional loss of the genes by mutation can be corrected by horizontal reacquisition, and naive organisms are more likely to acquire a clustered version than a collection of unclustered genes. The genes persist on a global scale because they can move horizontally fast enough to avoid elimination by insufficient (or episodic) selection.

The conventional (and reasonable) idea is that genes are maintained in a genome only if they provide a selectable phenotype. Selection continually purges a population of mutant individuals that have lost a valuable gene. The purging of mutants works best for genes that make a big fitness contribution (whose mutants are very deleterious). Conversely,

genes that are only weakly (or only occasionally) selected are most likely to be lost from a population by mutation and drift (35). Thus, all but the most essential genes are subject to loss by mutation in cells that are still encountered in the population. Genes that are weakly selected can enhance their survival by nonconventional (selfish) gimmicks, which become more important as their selective value drops.

From the point of view of a gene, the likelihood of persistence in the world is enhanced by horizontal transfer to new hosts, which provides a way of staying ahead of loss by mutation and counterselection (or drift). Transferred bacterial genes will be held in their new host only if they provide a phenotype of sufficient selective value. Many phenotypes require multiple proteins (and thus multiple genes). For a gene to take advantage of horizontal transfer as a way of maintaining itself, that gene must be transferred with all the other genes needed for its particular phenotype. This presents a problem in that only small DNA fragments are transferred from one bacterium to another. Thus a gene gains the advantage of horizontal transferability in direct proportion to the probability that it can be transferred simultaneously with all of the genes that allow it to confer a selectable phenotype—that is, in proportion to the proximity of the cooperating genes.

Two sets of genes, each contributing a single phenotype, may make the same fitness contribution regardless of their position in the genome. One set can gain the added selfish advantage of transferability if its genes are sufficiently close together to allow occasional cotransfer. The closer they are together, the more probable is this cotransfer. This provides a positive selection for clustering that becomes stronger as genes get closer together. Therefore, this model allows gene clustering to be achieved by a series of small steps; the selfish advantage increases progressively as the genes are brought closer together (36). Fusion of clustered genes into transcription units or operons is proposed to further enhance transmissibility by minimizing the problems of promoter adaptation.

This selfish operon model predicts that clustering will be a property of genes that are dispensable and subject to loss and reacquisition, and this is the general observation. For example, biosynthetic operons escape selection when their end product is present in the environment, as do genes for degradative pathways when cells grow on an alternative carbon and energy source. In contrast, clustering is not predicted for essential genes, which cannot be lost and thus might not be expected to be under selection for horizontal reacquisition. While essential genes are usually not clustered, there are exceptions. (The large operon encoding ribosomal proteins is especially notable.)

Most clusters of essential genes encode proteins that interact intimately. If a mutation arises that improves the function (e.g., protein synthesis), that new mutation would be under selection to spread globally if its gene could transfer horizontally. However, the mutant gene can move horizontally only if it brings along its companion proteins that work well together. The advantageous mutant protein may not work well when inserted in unfamiliar protein complexes. The transferred ribosomal gene clusters may supersede existing homologues rather than adding a novel function to a naive cell. Thus, intimate interactions between proteins may make it possible for horizontal transfer to drive clustering of essential genes. Because these clusters are essential, they are less prone to mutational loss or disruption. Evidence has been presented that clusters of this sort are stable over longer evolutionary periods (75).

A powerful tool has been developed based on the strong tendency of functionally related genes to cluster. While this has been implemented in several ways, a program we have found extremely powerful is the functional coupling program within the WIT (Integrated Genomics) program package (54). This program takes a submitted gene sequence and tabulates its neighbors (genes often found nearby in the chromosome) in a large number of complete bacterial genomes. The result is a list of genes that are frequently located near the submitted sequence. The ability of this program to identify known genes with related functions is astounding. This ability generates confidence that surprising examples of frequent neighbors are likely to reflect unexpected functional relatedness. In view of the selfish operon model, it seems likely that this powerful tool for identifying genes of related functions may owe its power to the selfish tendency of genes to enhance their survival by frequent horizontal transfer.

TRANSPOSONS AS SELFISH ELEMENTS

The above argument for selfish gene clusters is a less extreme example of an argument made long ago to explain the success of transposable elements. Insertion sequences (IS elements) are usually regarded as selfish elements that are dangerous to their hosts because they generate mutations and rearrangements that are almost all deleterious (see digression below). When considered as part of the bacterial genome, transposable elements maintain themselves despite counterselection, because replicative transposition allows them to stay ahead of purifying selection by

generating new copies faster than selection can remove them from the population. Replication can occur by transposition within the genome in an asexual lineage or by horizontal transfer to new lineages. By moving fast enough, these elements can stay ahead of counterselection without providing a beneficial function and can even overcome their substantial inherent cost in mutagenesis. They replicate more often per unit time than a conventional gene—a formula for success.

Some transposable elements contain drug resistance determinants that confer a selectable phenotype; this suggests that they might be maintained by conventional selection. This evidence for selective value does not contradict the initial idea that these elements are basically selfish. Multiple forms of selection can act simultaneously on any element. Transposons may depend primarily on selfish mechanisms for their survival, but enhance it by accumulating genes that provide a positively selectable function. This positive selection is demonstrable but insufficient to allow them to remain without their selfish behavior (see below). If the included genes could maintain themselves by conventional selection (e.g., for drug resistance), one would expect them to lose the costly ability to transpose and mutagenize the genome. Their continued association with transposition activity (and its deleterious effects on the host) suggests that the nonselfish aspects of their phenotype are not sufficient to ensure maintenance. It is proposed below that these nonselfish selective phenotypes are weak ones in a natural setting; i.e., drug resistance may only rarely be of selective importance.

Additional evidence suggesting a selective value of transposons has come from growth competition experiments in which strains with the elements do better than isogenic strains without them (10, 24). This advantage was seen for elements that encode no protein other than their own transposase. In the competition experiments, the advantage appeared to reflect a mutator effect in that the element caused mutations that allowed for improved growth (9, 23, 47). In the digression below, it is argued that these experiments reveal short-term advantages and do not assess the true long-term costs of carrying these elements. That is, the mutator phenotype is not likely to explain the persistence of transposable elements.

Transposons seem likely to be maintained primarily by a selfish advantage derived from transposition and horizontal transfer. In addition to being small enough to be transferred frequently, transposable elements actively contribute to their horizontal transfer by providing mechanisms for integration into the recipient genome; some also contribute to cell-cell transfer (63). Some of these elements enhance their maintenance by acquiring genes with selective value for the host, but the selfish behavior is still required for their maintenance. Transposable elements and some phages may also have "addictive" properties (see below).

A DIGRESSION ON THE COSTS OF MUTAGENESIS

While competition experiments suggest that transposable elements confer a selective advantage by their mutator effect, it seems more attractive that this is a short-term advantage seen under restricted conditions and is unlikely to explain the natural long-term persistence of these essentially selfish elements. The vast majority of random mutations are expected to be deleterious. This was shown directly for transposons (14, 62), but seems inescapable on first principles. There is a very large target for mutations that cause loss of valuable information (information maintained by selection). In contrast, there is likely to be a very tiny target for mutations that provide a fitness increase in any particular circumstance. For many conditions, the ratio of beneficial to deleterious mutations could easily approach 10^{-6} (20, 59). In short-term growth experiments, rare valuable mutations may arise and sweep the population selectively; adaptation may be aided by an increased mutation rate if the supply of mutations is limiting and if the fitness increase is large enough to outweigh the cost of the associated mutations. This can occur in small laboratory experiments that demand rare mutations with very large fitness increases (e.g., drug resistance) under a very specific set of conditions. Under these circumstances, mutation supply limits adaptation, and short-term benefit masks the long-term cost of the preponderant deleterious mutations.

Work done by Miller and coworkers (46) demonstrates clearly that mutators (strains with an increased mutation rate) are favored under conditions of strong selection (sequential acquisition of resistance to several drugs). This observation is often cited as evidence that increased mutation rate speeds genetic adaptation. The results demonstrate beautifully that mutators can be selected in short-term experiments, but do not address long-term cost. Later experiments showed that the cost of mutators is very high (18), making it unlikely that transposons are maintained in bacteria because they increase the mutation rate. There are easier ways to do this with mutation types that are more likely to be valuable, and it probably is not a good idea even then (59). The interacting parameters described by Muller (mutation

rate, population size, recombination frequency) can be set at a variety of compatible equilibrium values. However, once these values are set, they are likely to be addictive in the sense that increasing one parameter (e.g., mutation rate) can be costly. In general, mutagenesis brings a prohibitive long-term cost because deleterious mutations are so frequent. The numerical problem of increasing mutation rates has been discussed for one particular genetic system with pointers to some of the extensive previous literature (59).

PHAGE, COMPENSATION, AND ADDICTION

Phages are usually visualized as independent entities that happen to grow in bacteria. However, when viewed from a bacterial perspective (as an extension of the bacterial genome), they are the ultimate in selfish elements. They prosper by replicating and having a free-living form that provides for all aspects of horizontal transfer (exit, persistence outside the host, entrance and incorporation into the new genome). While virulent phages seem purely selfish, temperate phages clearly spend considerable time as part of a host chromosome. It is therefore not surprising that some temperate phages (like some transposable elements) have acquired genes that are not central to phage growth but confer a conventional selective advantage on their host (25). As for transposable elements, prophages appear to confer a selective advantage in mixed cultures, when lysogens and nonlysogens compete (7, 12, 37). This could reflect phage genes that provide a useful host phenotype. Regardless of these advantages, phages are likely to have strong costs to the bacterium in terms of lysis, and their persistence in the face of these costs suggests that they are substantially selfish. That is, their long-term maintenance still depends on occasional lytic growth (before mutational inactivation) and consequent horizontal transfer to new hosts. Survival of phage and transposons may also be enhanced by addiction mechanisms put in place by compensatory mutations in the host (see below).

Work on drug resistance has shown that when a strain becomes resistant (for example) to streptomycin by mutational ribosomal alteration, its resistance is associated with a growth disadvantage (in the absence of drug) due to impaired ribosome function (2). This impairment can be corrected (compensated) by secondary mutations that restore more rapid growth without loss of drug resistance (5). These compensatory mutations reduce fitness when present in strains without an initial streptomycin resistance mutation (44). It seems likely that many genetic situations may provide similar conditions. New mutations may provide a short-term benefit with a long-term cost. Compensatory mutations reduce the long-term cost without removing the benefit provided by the original mutation. However, if the original beneficial mutation is removed, the compensatory mutations impose a new cost. This phenomenon might contribute to some of the apparent selective benefit conferred by prophages and transposons.

If a wild-type bacterium (e.g., *E. coli* K-12 carrying phage lambda) spends considerable time growing with its lambda prophage, it may accumulate multiple mutations that minimize the cost of carrying that prophage. These mutations might, for example, minimize the probability of spontaneous prophage induction. When such a compensated strain is brought into the laboratory and cured of its prophage, the compensatory mutations remain and may be slightly deleterious in the absence of the prophage. Normal growth could be restored by replacement of the prophage and restoration of the conditions under which compensation was selected.

Similar events might explain growth advantages provided by certain transposable elements. That is, cells may have spent recent evolutionary time with such elements and acquired mutations that compensate for the cost of the transposon (e.g., reducing transposition frequency); such mutations may be slightly deleterious when the transposon is removed (especially under the rich conditions that force fast growth in the laboratory). Note that this argument suggests that some of the apparent selective value of phages and transposons could be due to host mutations that minimize the cost of carrying the element; the compensatory mutations serve to addict the host to its load of infective elements. This makes the elements appear valuable.

Many genetic features of a genome may be addiction modules in the sense that they do not contribute a useful phenotype but are tolerated better if certain compensatory genetic changes occur. (An example is the suggestion made above for addictive maintenance of gene order.) This raises serious problems for those trying to understand the physiological basis of the growth advantage provided by a phage or other genetic element. Such elements may in fact contribute nothing to normal cell function, even though a competitive disadvantage results from their removal.

Virulent phages (viewed as an extension of the bacterial genome) represent an extreme that may help to make these arguments clearer. From a bacterial point of view, these phages are solidly deleterious. One would not expect them to include genes that improve host fitness. Despite their deleterious effects, virulent phages persist because they can

replicate, accumulate mutations that circumvent resistance, and set up a standoff situation of balanced warfare in which both combatants (cells and phage) remain stably on battlefield. In essence, they are replicating and changing faster than they can be removed by selection. While it seems easier to view virulent phages as independent entities competing with bacteria, their underlying strategy is an extreme example of that used by temperate phages, transposons, and gene clusters: horizontal transfer is the fundamental underpinning of their survival.

THE PROBLEM OF PLASMIDS

Plasmids are important features of bacterial genomes that may represent a different aspect of selfish behavior. Conjugative plasmids (and mobilizable plasmids) are capable of intracellular transfer and thus are likely to be maintained in large part by selfish mechanisms (like phages and transposable elements), despite the fact that they often include genes that provide selectable cellular phenotypes. Nonconjugative plasmids present a problem.

Nonconjugative plasmids (in common with IS elements) are not obviously transmissible, but may show enhanced horizontal transfer because of their status as an independent replicon. This may enhance their ability to establish themselves in a new genome following transmission (by transduction or transformation); this is the same transmission benefit that transposition provides to an IS element. In this sense, nonconjugative plasmids may still be considered infectious selfish elements.

A second consideration in understanding plasmids is their high copy number. This might be considered an analogue of transposition (for IS elements), allowing genes to replicate faster than they can be removed by selection. However, multiple copies and genetic versions of a transposon can be held for extensive time in a single genome. High-copy-number plasmids (subject to a single-copy control mechanism) quickly clone themselves and do not stably maintain a heterozygous condition when different versions are carried by the same cell (53). Therefore, it seems likely that plasmid copy number may contribute a subtly different sort of selfish behavior; genes in high copy have a larger effective population size than chromosomal genes and thus may be maintainable by weaker selection (see Muller's rules above). In a sense, the higher copy number gives a measure of defense against mutation when an element is maintained by weak selective forces. Copy increase may be a selfish strategy that is distinct from infectious behavior.

STANDARD-LOOKING GENES CAN BENEFIT FROM SELFISH ASPECTS OF SELECTION

Selfish mechanisms used by phage, transposons, and plasmids can also contribute to maintenance of genes that are respectable chromosome residents. Below are examples of standard sequences with addictive or infectious behavior.

Restriction/modification systems are thought to confer a conventional selective advantage in that they protect against invading foreign (unmodified) DNA. However, the mutational loss of just the modification function is lethal if the restriction activity remains; this addicts cells (partially) to the modification function. While this may seem axiomatic, given the nature of the functions involved, it is interesting that simultaneous loss of both modification and restriction functions is also lethal for some systems, because the half-life of the existing restriction enzyme is longer than that of the modification enzyme (32, 51). It should be noted that modification genes can still be lost without cost if the restriction activity is removed first. This does not eliminate a contribution of selfish behavior to maintenance, because the situation still counterselects one means of functional loss.

Not all restriction enzyme systems show this addictive resistance to loss of both functions. This suggests that some bacterial lifestyles include a sufficient influx of harmful DNA that the restriction system is maintained solely by its conventional protective function. In contrast, environments that provide little influx of harmful DNA may reduce the selection on restriction systems sufficiently that they cannot be maintained in the genome by this mechanism alone. Under these conditions, a contribution from addictive behavior may be required for maintenance. The fact that *Helicobacter pylori* genomes include over 50 restriction modification systems (52) seems unlikely to reflect frequent exposure to invasive DNA (since these organisms inhabit sequestered sites) and may suggest (paradoxically) that the lifestyle of this bacterium includes so little infective DNA that only strongly addictive restriction systems have remained and that this addictive behavior has been sufficient to provide for both maintenance and horizontal spread.

An example of "infectious" behavior may be provided by the uptake signal sequences (USSs) found in *Haemophilus* and *Neisseria*. These sequences were originally considered part of a sexual exchange system that allows preferential uptake of conspecific DNA and is therefore highly overrepresented in the genome (70). However, it has been convincingly argued that bacteria take up DNA for use as a food source and that the concomitant sexual exchange is an unavoidable (and unintended) side effect (55). If

such a nutritional uptake system gained efficiency (lowered its K_m) by recognition of a specific sequence, it could take up extensive DNA sequences that include this sequence, even when the concentration of DNA is low. If this occurred, the genomic abundance of the recognized sequence might increase secondarily due to a selfish property—frequent horizontal transfer. The nature of the uptake system dictates that all incoming DNA sequences include the USS; therefore, to the extent that DNA is ever added to the chromosome from an outside source, the USS element copy number will increase. Stated differently, every time a USS arises by chance in a *Haemophilus* genome, it is likely to sweep all *Haemophilus* cells simply because it is efficiently taken up by all organisms that share the uptake mechanism and therefore has a chance of recombining into their genome with no required selection. In conclusion, a sequence element may be heavily overrepresented for selfish rather than functional reasons. Encountered without any knowledge of DNA transport, the overrepresented USS might well be a candidate for a role in chromosome packaging or movement, when in fact it may have evolved to improve uptake of a nutrient. A possible unifying idea would be that exogenous DNA is pulled into the cell following attachment of some multifunctional chromosome-moving mechanism at the specific USS.

USING GENE POSITION TO MAKE PRELIMINARY INFERENCES REGARDING SELECTIVE VALUE: A HIERARCHY OF SELFISHNESS

Above are described some genome features that depend to varying extents on their selfish behavior for their maintenance in the genome. These assumptions can be used in reverse to draw conclusions regarding the fitness contribution of genes based on their genomic position. The basic idea is the following: the more heavily an element depends on selfish behavior, the smaller is the likely contribution of its information to conventional fitness.

Consider a hierarchy of genes placed in order based on decreasing dependence on conventional selective value for their maintenance—and increasing dependence on selfish mechanisms (horizontal transmission or addiction). A shamelessly simplistic suggestion is shown in Table 1. At the top of the list are standard genes with important physiological roles and no need for selfish behavior. At the bottom are virulent phages that seem purely deleterious (from a bacterial point of view); they have no selective value and are maintained solely by their own selfish (infectious) means. Between these extremes is a series of elements ordered (very speculatively) according to perceived increasing dependence on selfish behavior. Ranking higher in the list than transmissible elements are standard chromosomal genes that are under conventional selection but enhance their (perhaps only moderate) selective value by having selfish properties that improve their horizontal transferability or cause a measure of addiction. Examples are functionally related gene clusters, genes flanked by repeated sequences, addictive genes, and sequences that enhance their own horizontal transfer by serving as transformation recognition signals (USSs mentioned above).

In the genome of one organism, a certain set of genes is unclustered and is therefore presumed to be maintained by conventional selection alone. In a different organism the same genes are driven to cluster, arguably by the added selfish benefit derived by enhanced horizontal transmissibility. The need for this extra measure of selection in the latter case suggests that these genes are of lower selective value to the second organism (in the case of a biosynthetic pathway, the end product may be more abundant in the environment of the second organism). The pathway is dispensable (at least temporarily) and its genes are therefore subject to mutational loss during periods of weak selection. Most clustered genes are dispensable (under some growth conditions) and are thus subject to mutational loss during periods of relaxed selection. Essential genes are rarely clustered. Exceptions to this (discussed above) are proposed to involve proteins that interact closely (36).

Genes with flanking repeats (described below) may enhance their chances of horizontal acquisition by having a means of excising by circularization and adding themselves to a new genome by homologous recombination following transfer. These repeat-flanked genes may also have an enhanced chance of amplifying during growth under selection (see below).

Next in Table 1 come conjugative plasmids and transposons that include conventionally selected genes but still maintain sophisticated mechanisms for infectious spread. These rank higher than simple IS elements, whose maintenance depends entirely on transposition. Temperate bacteriophages, which spend some time as part of the bacterial genome and may include genes with conventionally selectable functions, rank above IS elements with no included genes. These phages are maintained mainly by selfish transmissibility, but enhance their ability to remain in the genome by standard selection for included functions (and perhaps by host addiction). At the bottom of the hierarchy are elements that confer no fitness

Table 1. Genetic elements listed in order of decreasing fitness contribution (and increasing selfish behavior)

General behavior	Genetic feature	Means of selective maintenance	
		Conventional fitness contribution	Selfish behavior
Purely conventional selection	Essential chromosomal gene	High	None
	Chromosomal gene with important or frequent fitness contribution	Quite high	None
Conventional with slight contribution of selfish behavior	Clustered chromosomal genes of related function	Moderate	Clustering enhances horizontal transmissibility.
	Chromosomal genes flanked by repeats	Moderate	Flanking repeats aid in recombinational integration following transfer.
	Chromosomal genes with addictive property	Moderate	Addictive behavior gives resistance to mutational loss.
Mainly selfish with some conventional fitness contribution	Conjugative plasmid with a useful phenotype	Weak	Sophisticated machinery for horizontal transfer and maintenance as replicon
	Transposons conferring a selectable phenotype	Weak	Transposition ability allows escape from elimination and enhances inheritance following horizontal transfer.
	Temperate phages with host phenotypes	Very weak	Sophisticated mechanism for horizontal transfer and integration into genome of new host
Purely selfish	Insertion sequence without host phenotype	None or deleterious	Transposition minimizes loss by mutation and drift. Enhances integration following horizontal transfer.
	Plasmids without selectable phenotypes	None or deleterious	Controlled high copy number gives resistance to mutational loss. Replicon status is maintained after horizontal transfer.
	Temperate phages without phenotypes	None or deleterious	Sophisticated mechanism for horizontal transfer and integration into genome of new host
	Virulent phages	Deleterious	Possible addiction Rapid infectious transfer

gain on their host and may be neutral or deleterious. These elements depend entirely on selfish mechanisms to maintain a position in the genome.

This suggests a rule of thumb: the most strongly selected genes are likely to be located at standard positions in the chromosomes rather than clustered with genes contributing to the same function. The genes carried by highly developed selfish elements (plasmids, transposons, and phages) are probably under the weakest conventional selection. While some genes within a prophage, transposon, or plasmid may contribute a measurable fitness increase, one does not expect a cell's important or essential genes to be found there. In considering the validity of this suggestion, it is important to remember that a weakly selected gene may either make a continuous very small contribution to fitness or, alternatively, could be absolutely essential under conditions that are only rarely encountered (e.g., drug resistance).

DO BACTERIAL CHROMOSOMES EXPLOIT A "WIND OF SUPERCOILING"?

The replication origin and terminus of *E. coli* define the ends of two equal-sized domains replicated divergently from a single origin. This feature is combined with a preferential orientation of the most heavily transcribed genes and operons so that their transcription proceeds in the same direction as their replication. It seems attractive to imagine that these two features reflect interacting effects of supercoiling. Both transcription and replication introduce positive supercoiling ahead of their direction of travel and

negative supercoils in their wake. The positive supercoiling introduced by one terminus-directed transcription unit would be taken up by the next (more terminus-proximal) promoter oriented in the same direction. Thus, heavy transcripts would avoid interfering with each other. Similarly, a replication fork would be expected to inhibit activity of promoters ahead of its path and stimulate those it has passed. One might imagine that supercoils introduced in this way are pushed inexorably toward the terminus, where serious problems seem likely to arise.

THE PROBLEM OF CONVERGING REPLICATION FORKS

While minor local supercoiling problems might be solved by topoisomerases, it seems possible that something special is needed to solve the more intense problems that arise as replication forks converge. Each fork introduces positive coils ahead of its path, and these are focused on a smaller and smaller region as the forks come closer together. If it becomes difficult to open DNA (or to relieve this supercoiling), one might expect a blockage of fork movement. Solving this problem might require some special features of the termination region (e.g., action of XerCD, temporary strand interruptions at or near *dif*) or alternatively by ends of linear chromosomes (below).

To the extent that fork convergence requires specialized sequence features present in the terminus region, it poses a problem for sister-strand recombination events. It seems increasingly attractive to imagine that recombination involves initiation of new replication forks (38). Two replication forks generated by recombinational repair of a double-strand break could converge in a region not designed to accommodate such forks. Ter and Chi sites may contribute to a system for minimizing this problem (see below).

CIRCULAR VERSUS LINEAR CHROMOSOMES

The form of a chromosome seems profoundly important to many models for compaction and movement. Yet many bacteria support both linear and circular replicons, suggesting that the difference is not as profound as one might imagine (8). Despite the difference in chromosome structure, however, the distribution and orientation of genes (and the consequent skew in base composition) is biased in *Borrelia*, just as it is in *E. coli* (19; also see below). It will be interesting to see if the *E. coli* chromosome can segregate normally if its circular chromosome is disrupted (perhaps at the *dif* site) and the two ends are supplied with the sort of telomeres found in chromosomes and plasmids of *Borrelia*. The switch might be managed by the circle-linear interconverting function found in the lambdoid phage N15 (33). It seems attractive to imagine that the difference between linearity and circularity is not central to the behavior of bacterial chromosomes.

The supercoiling problem associated with converging replication forks near the terminus of a circular chromosome, discussed above, seems to be solved for linear chromosomes, whose ends might be expected to rotate freely. Compaction of bacterial chromosomes by supercoiling might be harder to visualize for linear chromosomes if their ends are free to rotate. However, if supercoiled domains are defined as regions between points at which the chromosome is either cross-linked or anchored (perhaps to the membrane), then the difference between linear and circular structures seems less critical.

SHORT ELEMENTS THAT SEEM IMPORTANT TO CHROMOSOME BEHAVIOR: Ter AND Chi

Two short sequence elements have been extensively studied with regard to their effects on chromosome replication (Ter) and recombination (Chi). Replication forks terminate when they encounter a properly oriented Ter sequence with a bound Tus protein (27). Following chromosome breakage, degradation of a double-strand end by RecBCD is terminated at a properly oriented Chi sequence, producing a single-strand 3′ end that can be used by RecA to initiate recombination, possibly by reestablishing a replication fork (34, 69, 73). While Ter and Chi have been extensively analyzed mechanistically, their biological roles seem less clear. Both elements show a nonrandom distribution and orientation in the genome of *E. coli* and *Salmonella*.

Ter sequences are located in the half of the chromosome farthest from the origin and are symmetrically distributed around the *dif* site, where replication normally terminates and where the XerCD functions act to ensure partition of completed chromosome copies to daughter cells. These Ter sites are oriented so as to allow passage of forks headed toward *dif* and to block rare replication forks that manage to pass the *dif* site. It has been proposed that Ter is a "fail-safe" device to ensure that opposed replication forks meet and are terminated near *dif*; yet some Ter sites are located far from *dif* and would only affect forks that had passed multiple similarly oriented Ter sites.

Chi sequences are overrepresented in all parts of the genome and are preferentially oriented such that RecBCD acting at a double-strand break would favor recombinational activation of the end nearest the origin and degradation of the end nearest the terminus; 75% of Chi sequences are in this orientation. The reasons for this are unclear.

Perhaps Chi and Ter elements are parts of a system that acts to ensure that replication forks formed during DNA repair recombination are directed toward the terminus region (*dif*) and not toward each other or toward new forks progressing from the origin. This may help avoid the supercoiling problem generated when forks converge, especially when they converge in regions that are not equipped to deal with the problem (Fig. 1; also see below).

At a double-strand break, the two ends are subject to resection and can then invade a sister chromosome to initiate forks. Depending on the timing of these invasions, forks that diverge or converge could be produced, or one fork could be lost (Fig. 1). The distribution of Chi sequences would favor prompt activation of the end nearest the origin and extensive degradation of the end nearest the terminus. If degradation proceeds to the old fork, the origin-proximal fork (directed toward the terminus) supersedes the old fork; this result is favored by the action of Chi and RecBCD. If the terminus-proximal end (following degradation) encounters a rare reversed Chi and initiates a fork, that fork is likely to converge with the origin-proximal fork and, if problems result, might have a chance to abort and try again. The third possibility is that one of the forks starts quickly and passes the other free end before it can initiate a fork, resulting in divergent forks, with one of them opposed to any new fork coming from the origin. An important function of Ter may be to terminate these "escaped" recombinational forks. Thus, the combination of Chi and Ter systems may act to preferentially direct new recombinational forks toward the terminus. Ter sites may be less essential near the origin, since forks that escape in this region can proceed across the origin.

THE GENOME OF *SALMONELLA ENTERICA* SEROVAR TYPHI: A TEST CASE FOR MAINTAINING GENE ORDER

It was suggested above that chromosomal gene order is under selection (perhaps by conventional functional selection to maintain a wind of supercoiling, perhaps by some selfish mechanism). It is striking that the basic gene order in *S. enterica* and *E. coli* has been maintained for perhaps 200 million years—the aggregate time elapsed as each diverged from their common ancestor. However, isolates of *S. enterica* serovar Typhi (40) and other host-specific

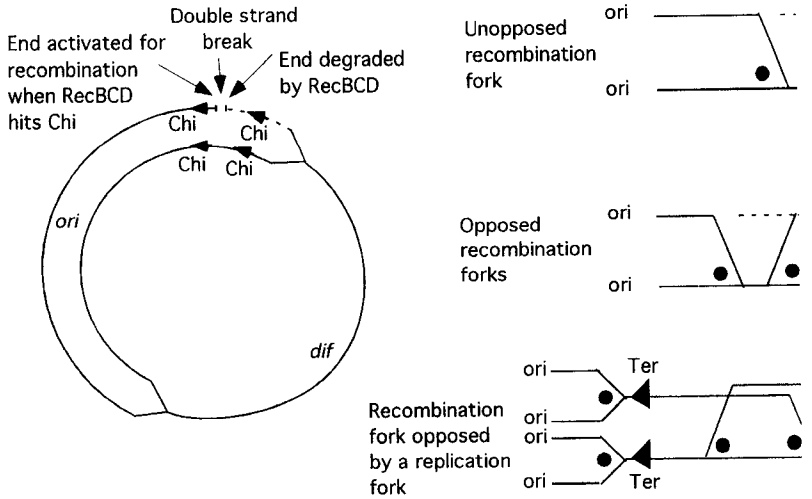

Figure 1. A proposal for the role of Chi and Ter sites in managing recombinational replication. (Left) Diagram of a double-strand break and the positions of nearby Chi sequences in their predominant orientation. The left end is likely to be promptly activated for recombination and lead to a fork moving toward the terminus. The right end is likely to be extensively degraded before engaging in recombination and could form a fork moving toward the origin. (Right) Diagram of the possible outcomes if the two broken ends establish independent recombinational forks. Depending on the timing of fork initiation, the two forks may diverge, or converge, or one end may be destroyed, leaving a single fork. Forks bound toward the origin might be terminated at Ter sites before converging with approaching replication forks. The concerted action of Chi and Ter might act to minimize fork collision and preferentially direct all recombinational replication toward the terminus.

Salmonella strains (39) provide exceptions that may help test rules for chromosome conservation.

In *E. coli* and *S. enterica*, the positions and orientation of seven repeated rRNA operons (*rrn*) are conserved. There are two widely separated, direct-order copies on one side of the origin and five similarly separated direct-order copies on the other side; all are oriented so that transcription proceeds away from the origin of replication. Although these extensive repeated sequences provide targets for recombination that could lead to inversions and transpositions of blocks of sequence, these rearrangements are not found in most natural isolates. In contrast, isolates of *S. enterica* serovar Typhi (a human-specific pathogen) show rearrangements that result from exchanges between these prominent *rrn* repeats. The behavior of serovar Typhi does not seem to indicate that the organism is more proficient at recombination between internal repeats; inversions form readily between inverse-order repeats in *E. coli* and in *S. enterica* (13).

The serovar Typhi rearrangements do not disturb the "wind of supercoiling" or the bias in GC content (described below). If all rearrangements are achieved by recombination between *rrn* repeats, none of the resulting rearrangements will reverse the orientation of any gene vis-à-vis the direction of replication. This is a necessary consequence of the fact that all *rrn* loci within a single replicore are located in the same orientation. That is, inversions made by recombination between *rrn* sequences must all include the origin (or terminus) of replication. These rearrangements could, however, alter the distance between a particular gene and the replication origin, but most observed rearrangements seem to conserve this spacing. This may suggest that the observed rearrangements are not random, but conserve the features that are most critical to chromosome function. Alternatively, serovar Typhi may grow under conditions (or at a rate) that relax some of the selective constraints on gene order. The rearrangements observed in serovar Typhi may be mechanistically permitted for all enteric bacteria, but the frequency distribution of permitted arrangements is broader (thus the selective cost of rearrangement is lower) in this particular subgroup.

INTERPRETING SMALL SEQUENCE ELEMENTS THAT ARE OVERREPRESENTED IN THE CHROMOSOME

Several small sequence elements (not obviously transposable) are found in the genome of *E. coli* (31). The REP (or PU) elements are imperfect palindromes, 30 to 40 bp in length (21, 26, 72), that have been placed in several general sequence classes (3, 22). These elements appear frequently at intergenic sites within transcribed regions as arrays in direct or inverse order. The palindromic character of these elements appears better conserved toward the outside ends, reminiscent of some transposable elements. While these sequences are frequently discussed as remnants of transposons, they have never been shown to transpose, and the conservation of their sequences suggests strongly that they may now be under selection and may play some role in cell physiology (or may be maintained by some unrecognized selfish mechanism).

Evidence that REP elements modulate the relative amounts of proteins encoded in a single message suggested a role in message half-life or translation (72). However, some of these units are located outside regions known to be transcribed, and evidence has been presented for interaction between these sequences and DNA-binding enzymes, including gyrase (76), IHF factor (6), and PolA. One of the largest effects of REP elements is their participation in formation of duplications and deletions (68). It seems possible that these REP sequences will prove important to bacterial chromosome behavior.

ERIC or IRU elements are larger and have a more complex secondary structure (28). While functions have not been demonstrated for these elements, they may stimulate recombination and be sites of frequent replication fork stalling or chromosome breakage (67; R. Dawson and J. R. Roth, unpublished).

STRAND-BIASED BASE SUBSTITUTION RATES

The base composition of the two strands of many bacterial chromosomes is skewed (the two strands show different overall base compositions). This is most striking for G/C pairs, which are positioned such that an excess of G residues appears in the leading strand (the untranscribed, sense strand for most highly expressed genes, which use the lagging strand as transcription template). Thus, this skew shows striking reversal at the origin and terminus of replication (42, 43). Evidence has been presented that this skew is the result of differential mutational pressure on the two strands.

Much of this bias is now thought to result from more frequent cytosine deamination (C → T changes) in single-stranded versus double-stranded DNA. It is suggested that transcription displaces the nontemplate strand and thereby exposes its C residues to deamination; over time this reduces the C level and increases the T level in this strand. This becomes

a genome-wide bias because heavily expressed genes tend to be oriented so their transcription proceeds away from the origin; i.e., their transcription template is the lagging strand (17). In principle, this strand bias should reverse for each gene in the opposite orientation. However, the observed strand bias is heavier for genes transcribed away from the origin. This may reflect the fact that genes transcribed away from the origin are more heavily transcribed and consequently subject the leading strand to more frequent deamination of C residues; alternatively, the leading strand may be subject to superimposed additional deamination because it is exposed as a single strand when it serves as template for lagging strand replication (56).

Regardless of the underlying reasons for strand bias in base composition, the leading strand is generally richer in G residues (and slightly richer in T residues) than the lagging strand in many bacteria (45, 57). Because of this bias, any G-rich sequence element is more likely to appear by chance in the leading strand and will show a strand preference that reverses sharply at the terminus and origin. This systematic skew is seen in many bacterial genomes and has been used to infer origins and termini of replication.

ORIENTATION-BIASED OLIGONUCLEOTIDE SEQUENCES

Inspection of genome sequences has revealed short, nonpalindromic oligonucleotides that are preferentially oriented vis-à-vis the direction of chromosome replication (60). The abundance of some of these sequences increases in the region of the terminus, and their orientation preference shifts sharply at the *dif* site. These elements (perhaps justifiably) have irresistible appeal for those investigating the nature of the terminus or looking for the attachment sites of proteins responsible for chromosome movement. However, the considerations in the preceding section predict that G-rich motifs will be overrepresented in the leading strand simply because of the higher G content of this strand. Thus, overrepresented or strand-biased sequences being considered for a role in cell physiology must be well analyzed to be sure they reflect more than secondary consequences of mutationally generated skew in base composition (see above). Increasing frequency near the terminus may not be sufficient, since mutation rates appear to increase with distance from the origin (66). Two such G-rich elements (Chi and Rag) show a directional bias (75 and 82% in the leading strand) that is clearly in excess of that predicted by GC skew for sequences of the same composition (56%) (41).

CONVENTIONALLY SELECTED SEQUENCE ELEMENTS CAN HAVE FUNCTIONS THAT ARE SEEN ONLY AT THE POPULATION LEVEL

There is a formal possibility that some conventionally maintained (nonselfish) sequence elements have functions that can only be understood at the population level. For example, in *Haemophilus* and *Neisseria*, base runs appear in genes (contingency loci) that encode proteins involved in the interactions between the bacterium and its host (4). At these particular sequences, there is a high probability of frameshift mutations that inactivate (and reactivate) the gene, making a sort of on-off toggle switch. Each of the multiple contingency loci is thereby switched on and off stochastically, generating a variety of individuals in the population with different patterns of gene expression. The basic idea is that any appreciable population will include at least one individual that will show a pattern that allows it to face the latest challenge thrown up by the host. This behavior should be distinguished from that of mutators (discussed above), whose high frequency of randomly distributed deleterious mutations seems unlikely to be beneficial (59). In the case of contingency loci, the mutations affect a limited gene set, and the (on-off) reversibility prevents long-term destruction of information.

Another example might be directly repeated sequence elements flanking genes or including genes whose amplification is frequently selected (16, 58, 71). These repeats would have no immediate function within a cell, but might enhance the frequency of a particular duplication and allow amplification under selection. The role of these repeats would be difficult to analyze functionally using conventional genetic analysis. This sort of mechanism has been suggested for genes involved in nitrogen fixation which are repeated in a plasmid. These repeats support amplification of the genes for nitrogen fixation when the cells are placed under selective conditions. Repeats of this sort might be thought of as contributing to programmed amplification or as parts of a regulatory mechanism whose specificity is provided by natural selection.

In both contingency loci and the hypothetical programmed amplification, the selected event is reversible and the conserved sequence elements support interconversion of a few states for a small number of particular genes. In this sense, the examples resemble phase variation mechanisms, which might be regarded as another case in which identifiable sequence elements (genes and noncoding sequences) have a biological function that can be understood only at a population level.

SUMMARY

Whole chromosomes show us all the information that they carry. Some of this information is highly functional and intensely informative for those seeking a mechanistic understanding of chromosome behavior and cell function. However, much information owes its presence to events that occur at a population level and are unlikely to contribute to single-cell physiology. These features include selfish elements and conventional sequences whose importance is realized only in a population. In looking for sequence elements that are directly relevant to chromosome packaging, movement, and expression, the selfish and population-based functions may be a fascinating distraction.

Acknowledgments. I thank Dan Andersson, Dan Dykhuizen, Glenn Herrick, Pat Higgins, Ichizo Kobayashi, Jeff Lawrence, Jon Seger, Valley Stewart, and David Witherspoon for stimulating discussions and useful suggestions. They are not responsible for any inadequacies in presentation or errors in fact.

REFERENCES

1. **Ames, B. N., and R. G. Martin.** 1964. Biochemical aspects of genetics: the operon. *Annu. Rev. Biochem.* 33:235–258.
2. **Andersson, D. I., and B. R. Levin.** 1999. The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* 2:489–493.
3. **Bachellier, S., J. M. Clement, M. Hofnung, and E. Gilson.** 1997. Bacterial interspersed mosaic elements (BIMEs) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence. *Genetics* 145:551–662.
4. **Bayliss, C. D., D. Field, and E. R. Moxon.** 2001. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J. Clin. Investig.* 107:657–662.
5. **Björkman, J., I. Nagaev, O. G. Berg, D. Hughes, and D. I. Andersson.** 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science* 287:1479–1482.
6. **Boccard, F., and P. Prentki.** 1993. Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3′ end of transcription units. *EMBO J.* 12:5019–5027.
7. **Campbell, J. H., D. Dykhuizen, and B. G. Rolfe.** 1978. Effects of the rex gene of phage lambda on lysogeny. *Genet. Res.* 32:257–263.
8. **Casjens, S., N. Palmer, R. van Vugt, W. M. Huang, B. Stevenson, P. Rosa, R. Lathigra, G. Sutton, J. Peterson, R. J. Dodson, D. Haft, E. Hickey, M. Gwinn, O. White, and C. M. Fraser.** 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.* 35:490–516.
9. **Chao, L., and S. M. McBroom.** 1985. Evolution of transposable elements: an IS*10* insertion increases fitness in *Escherichia coli*. *Mol. Biol. Evol.* 2:359–369.
10. **Chao, L., C. Vargas, B. B. Spear, and E. C. Cox.** 1983. Transposable elements as mutator genes in evolution. *Nature* 303:633–635.
11. **Demerec, M., and P. Hartman.** 1959. Complex loci in microorganisms. *Annu. Rev. Microbiol.* 13:377–406.
12. **Edlin, G., L. Lin, and R. Bitner.** 1977. Reproductive fitness of P1, P2, and Mu lysogens of *Escherichia coli*. *J. Virol.* 21:560–564.
13. **Edwards, R. A., G. J. Olsen, and S. R. Maloy.** 2002. Comparative genomics of closely related Salmonellae. *Trends Microbiol.* 10:94–99.
14. **Elena, S. F., L. Ekunwe, N. Hajela, S. A. Oden, and R. E. Lenski.** 1998. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* 102-103:349–358.
15. **Felsenstein, J.** 1974. The evolutionary advantage of recombination. *Genetics* 78:737–756.
16. **Flores, M., P. Mavingui, X. Perret, W. J. Broughton, D. Romero, G. Hernandez, G. Davila, and R. Palacios.** 2000. Prediction, identification, and artificial selection of DNA rearrangements in Rhizobium: toward a natural genomic design. *Proc. Natl. Acad. Sci. USA* 97:9138–9143.
17. **Francino, M. P., L. Chao, M. A. Riley, and H. Ochman.** 1996. Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science* 272:107–109.
18. **Funchain, P., A. Yeung, J. L. Stewart, R. Lin, M. M. Slupska, and J. H. Miller.** 2000. The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* 154:959–970.
19. **Garcia-Lara, J., M. Picardeau, B. J. Hinnebusch, W. M. Huang, and S. Casjens.** 2000. The role of genomics in approaching the study of *Borrelia* DNA replication. *J. Mol. Microbiol. Biotechnol.* 2:447–454.
20. **Gerrish, P. J., and R. E. Lenski.** 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102-103:127–144.
21. **Gilson, E., J. M. Clement, D. Brutlag, and M. Hofnung.** 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J.* 3:1417–1421.
22. **Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung.** 1991. The BIME family of bacterial highly repetitive sequences. *Res. Microbiol.* 142:217–222.
23. **Hartl, D. L., D. E. Dykhuizen, and D. E. Berg.** 1984. Accessory DNAs in the bacterial gene pool: playground for coevolution. *Ciba Found. Symp.* 102:233–245.
24. **Hartl, D. L., D. E. Dykhuizen, R. D. Miller, L. Green, and J. de Framond.** 1983. Transposable element IS*50* improves growth rate of *E. coli* cells without transposition. *Cell* 35:503–510.
25. **Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens.** 2000. The origins and ongoing evolution of viruses. *Trends Microbiol.* 8:504–508.
26. **Higgins, C. F., G. F. Ames, W. M. Barnes, J. M. Clement, and M. Hofnung.** 1982. A novel intercistronic regulatory element of prokaryotic operons. *Nature* 298:760–762.
27. **Hill, T. M.** 1996. Features of the chromosomal terminus region, p. 1602–1614. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, 2nd ed., vol. 2. ASM Press, Washington, D.C.
28. **Hulton, C. S., C. F. Higgins, and P. M. Sharp.** 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* 5:825–834.
29. **Jacob, F., and J. Monod.** 1962. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* 26:193–211.
30. **Jacob, F., D. Perrin, C. Sanchez, and J. Monod.** 1960. L'opéron: groupe de gènes à expression coordinée par un opérateur. *Crit. Rev. Acad. Sci.* 250:1727–1729.

31. **Kawano, M., S. Kanaya, T. Oshima, Y. Masuda, T. Ara, and H. Mori.** 2002. Distribution of repetitive sequences on the leading and lagging strands of the *Escherichia coli* genome: comparative study of Long Direct Repeat (LDR) sequences. *DNA Res.* **9**:1–10.
32. **Kobayashi, I.** 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* **29**:3742–3756.
33. **Kobryn, K., and G. Chaconas.** 2001. The circle is broken: telomere resolution in linear replicons. *Curr. Opin. Microbiol.* **4**:558–564.
34. **Kowalczykowski, S. C.** 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**:156–165.
35. **Lawrence, J. G., and J. R. Roth.** 1999. Genomic flux: genome evolution by gene loss and acquisition, p. 263–289. *In* R. L. Charlebois (ed.), *Organization of the Prokaryotic Genome.* ASM Press, Washington, D.C.
36. **Lawrence, J. G., and J. R. Roth.** 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**:1843–1860.
37. **Lin, L., R. Bitner, and G. Edlin.** 1977. Increased reproductive fitness of *Escherichia coli* lambda lysogens. *J. Virol.* **21**:554–559.
38. **Liu, J., L. Xu, S. J. Sandler, and K. J. Marians.** 1999. Replication fork assembly at recombination intermediates is required for bacterial growth. *Proc. Natl. Acad. Sci. USA* **96**:3552–3555.
39. **Liu, S. L., and K. E. Sanderson.** 1998. Homologous recombination between *rrn* operons rearranges the chromosome in host-specialized species of *Salmonella. FEMS Microbiol. Lett.* **164**:275–281.
40. **Liu, S. L., and K. E. Sanderson.** 1995. Rearrangements in the genome of the bacterium *Salmonella typhi. Proc. Natl. Acad. Sci. USA* **92**:1018–1022.
41. **Lobry, C., and J. Louarn. 2003.** Polarization of prokaryotic chromosomes. *Curr. Opin. Microbiol.* **6**:101–108.
42. **Lobry, J. R.** 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**:660–665.
43. **Lobry, J. R., and C. Lobry.** 1999. Evolution of DNA base composition under no-strand-bias conditions when the substitution rates are not constant. *Mol. Biol. Evol.* **16**:719–723.
44. **Maisnier-Patin, S., O. G. Berg, L. Liljas, and D. I. Andersson.** 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium. Mol. Microbiol.* **46**:355–366.
45. **McLean, M. J., K. H. Wolfe, and K. M. Devine.** 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol.* **47**:691–696.
46. **Miller, J. H., A. Suthar, J. Tai, A. Yeung, C. Truong, and J. L. Stewart.** 1999. Direct selection for mutators in *Escherichia coli. J. Bacteriol.* **181**:1576–1584.
47. **Miller, R. D., D. E. Dykhuizen, and D. L. Hartl.** 1988. Fitness effects of a deletion mutation increasing transcription of the 6-phosphogluconate dehydrogenase gene in *Escherichia coli. Mol. Biol. Evol.* **5**:691–703.
48. **Muller, H.** 1932. Some genetic aspects of sex. *Am. Nat.* **66**:118–138.
49. **Muller, H. J.** 1964. The relation of recombination to mutational advance. *Mutat. Res.* **1**:2–9.
50. **Mushegian, A., and E. Koonin.** 1996. Gene order is not conserved in bacterial evolution. *Trends Genet.* **12**:289–290.
51. **Naito, T., K. Kusano, and I. Kobayashi.** 1995. Selfish behavior of restriction-modification systems. *Science* **267**:897–899.
52. **Nobusato, A., I. Uchiyama, and I. Kobayashi.** 2000. Diversity of restriction-modification gene homologues in *Helicobacter pylori. Gene* **259**:89–98.
53. **Novick, R. P., and F. C. Hoppensteadt.** 1978. On plasmid incompatibility. *Plasmid* **1**:421–434.
54. **Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev.** 1999. Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol.* **1**:93–108.
55. **Redfield, R. J.** 2001. Do bacteria have sex? *Nat. Rev. Genet.* **2**:634–639.
56. **Rocha, E. P., and A. Danchin.** 2001. Ongoing evolution of strand composition in bacterial genomes. *Mol. Biol. Evol.* **18**:1789–1799.
57. **Rocha, E. P., A. Danchin, and A. Viari.** 1999. Universal replication biases in bacteria. *Mol. Microbiol.* **32**:11–16.
58. **Romero, D., and R. Palacios.** 1997. Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* **31**:91–111.
59. **Roth, J. R., E. Kofoid, F. P. Roth, O. G. Berg, J. Seger, and D. I. Andersson.** 2003. Regulating general mutation rates: examination of the hypermutable state model for Cairnsian adaptive mutation. *Genetics* **163**:1483–1496.
60. **Salzberg, S. L., A. J. Salzberg, A. R. Kerlavage, and J. F. Tomb.** 1998. Skewed oligomers and origins of replication. *Gene* **217**:57–67.
61. **Schmid, M. B., and J. R. Roth.** 1987. Gene location affects expression level in *Salmonella typhimurium. J. Bacteriol.* **169**:2872–2875.
62. **Schneider, D., E. Duperchy, E. Coursange, R. E. Lenski, and M. Blot.** 2000. Long-term experimental evolution in *Escherichia coli.* IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* **156**:477–488.
63. **Scott, J. R., and G. G. Churchward.** 1995. Conjugative transposition. *Annu. Rev. Microbiol.* **49**:367–397.
64. **Sharp, P. M., and W. H. Li.** 1987. The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**:1281–1295.
65. **Sharp, P. M., and W. H. Li.** 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**:28–38.
66. **Sharp, P. M., D. C. Shields, K. H. Wolfe, and W. H. Li.** 1989. Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* **246**:808–810.
67. **Sharples, G. J., and R. G. Lloyd.** 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* **18**:6503–6508.
68. **Shyamala, V., E. Schneider, and G. F. Ames.** 1990. Tandem chromosomal duplications: role of REP sequences in the recombination event at the join-point. *EMBO J.* **9**:939–946.
69. **Smith, G. R., S. M. Kunes, D. W. Schultz, A. Taylor, and K. L. Triman.** 1981. Structure of chi hotspots of generalized recombination. *Cell* **24**:429–436.
70. **Smith, H. O., M. L. Gwinn, and S. L. Salzberg.** 1999. DNA uptake signal sequences in naturally transformable bacteria. *Res. Microbiol.* **150**:603–616.
71. **Sonti, R., and J. R. Roth.** 1988. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth limiting carbon sources. *Genetics* **123**:19–28.
72. **Stern, M. J., G. F. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37**:1015–1026.
73. **Thaler, D. S., and F. W. Stahl.** 1988. DNA double-chain breaks in recombination of phage lambda and of yeast. *Annu. Rev. Genet.* **22**:169–197.

74. **Wheelis, M., and R. Stanier.** 1970. The genetic control of dissimilatory pathways in *Pseudomonas putida*. *Genetics* 66:245–266.
75. **Wolf, Y. I., I. B. Rogozin, A. S. Kondrashov, and E. V. Koonin.** 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res.* **11**:356–372.
76. **Yang, Y., and G. F. Ames.** 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc. Natl. Acad. Sci. USA* 85:8850–8854.