ANNUAL REVIEWS **Further**

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

The Organization of the Bacterial Genome

Eduardo P.C. Rocha

Institut Pasteur, Microbial Evolutionary Genomics, CNRS, URA2171, F-75015 Paris, France; email: erocha@pasteur.fr

Annu. Rev. Genet. 2008. 42:211–33

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev.genet.42.110807.091653

Copyright © 2008 by Annual Reviews. All rights reserved

0066-4197/08/1201-0211$20.00

Key Words

replication, transcription, nucleoid, segregation, rearrangements, evolution

Abstract

Many bacterial cellular processes interact intimately with the chromosome. Such interplay is the major driving force of genome structure or organization. Interactions take place at different scales—local for gene expression, global for replication—and lead to the differentiation of the chromosome into organizational units such as operons, replichores, or macrodomains. These processes are intermingled in the cell and create complex higher-level organizational features that are adaptive because they favor the interplay between the processes. The surprising result of selection for genome organization is that gene repertoires change much more quickly than chromosomal structure. Comparative genomics and experimental genomic manipulations are untangling the different cellular and evolutionary mechanisms causing such resilience to change. Since organization results from cellular processes, a better understanding of chromosome organization will help unravel the underlying cellular processes and their diversity.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

INTRODUCTION

HGT: horizontal gene transfer

After the publication of hundreds of complete prokaryotic genomes few would underestimate the role of genomics in contemporary molecular microbiology. DNA sequencing facilitates genetic manipulation and promises to uncover the basic functional schemas of the uncultivable microbial majority. Genome data have also highlighted the very peculiar mode of genome evolution in prokaryotes when compared to model eukaryotes. The genomes of the latter evolve new functions mostly by gene duplication; their substrates, chromosomes, have very distinctive regions, notably centromeres and telomeres. Their transcription units usually include one single gene, and their cellular processes are highly compartmentalized. In prokaryotes, the gene repertoires increase mostly by horizontal gene transfer (HGT), not by duplication. Chromosomes are relatively uniform in terms of gene density and sequence composition. Genes are typically cotranscribed in operons. Many cellular processes are coupled. Whereas two strains of *Escherichia coli* have more unrelated genes than two typical mammalian genomes, genome maps of *E. coli* and *Bacillus subtilis*, which diverged several billion years ago, are more similar than are yeast genomes diverged a few hundred millions years ago. As a consequence, bacterial chromosomes tend to have architectures that are both complex and plastic, albeit generally very different from eukaryotes. Although they are surprisingly flexible in terms of gene repertoires, their organizational features are highly conservative (**Figure 1**).

In this review, I argue that available evidence shows that all cellular processes interacting directly or indirectly with DNA affect and shape genome structure. The underlying molecular cause is that such processes impose constraints and/or lead to selection of some favorable configurations of genomic objects. Naturally, if two processes interact in the chromosome then the affected regions will be constrained by the processes and their interaction, which requires fine-tuned organization. The resulting picture

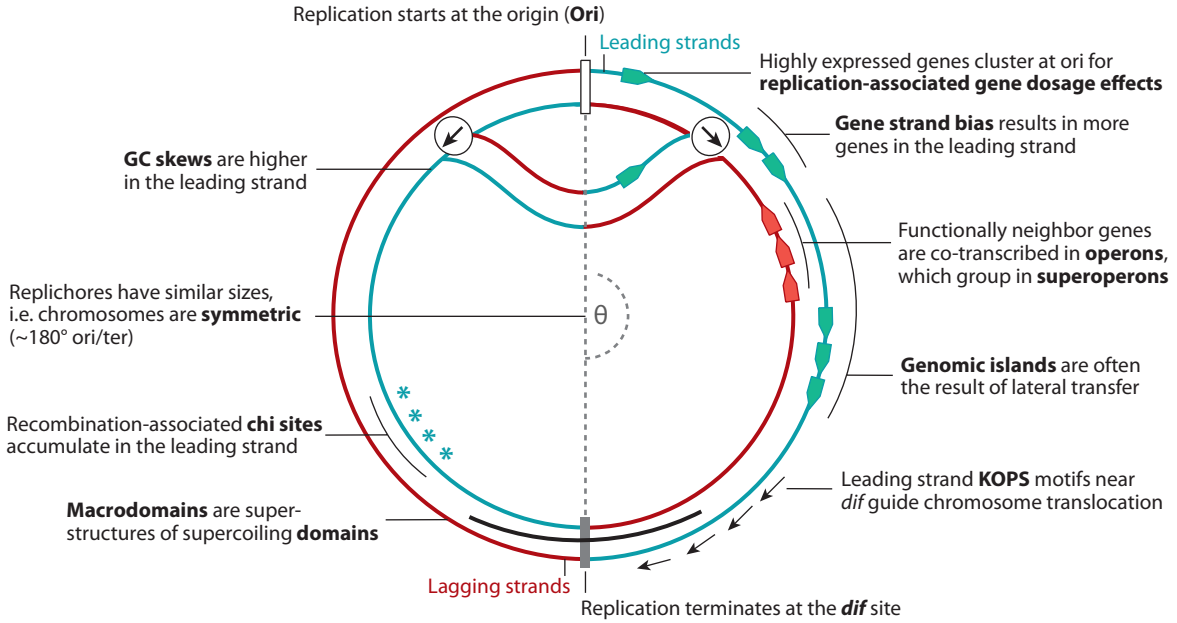


Figure 1

Elements of genome organization.

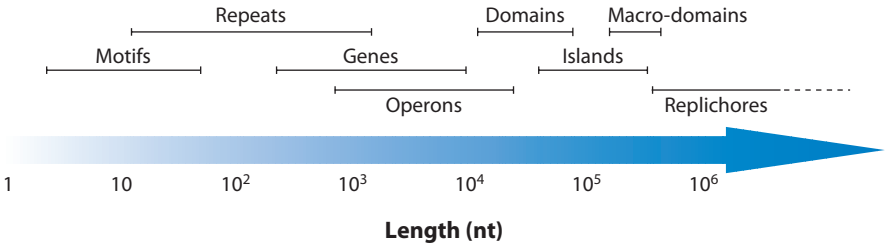


Figure 2

The scales of genome organization.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

is that at the crossroads between interactions genomes become highly organized. Each section of this review is thus titled after one or several cellular processes whose interplay shapes chromosome organization. Examples of such emerging organizational features include the overabundance of leading strand genes caused by the antagonistic interaction between replication and gene expression, the biases in gene distribution that may favor chromosome segregation by way of gene expression, or the aggregation of functionally neighboring operons to benefit from the effects of nucleoid opening in coexpression. These features shape chromosome organization at very different scales, from small motifs to very large chromosomal regions (**Figure 2**). Since the organization of genetic information is adaptive, most spontaneous rearrangements lead to lower fitness, e.g., by slowing growth. The conflict between genome dynamics and chromosome organization is molded by natural selection and depends on ecological and cellular processes that may thus be unraveled by comparative genomics.

GENE EXPRESSION AND CHROMOSOME COLOCALIZATION

Half a century ago it became apparent that related enzyme-coding genes tend to be colocalized in the bacterial chromosome. Furthermore, the order of these genes follows the order of the corresponding enzyme activities in metabolic pathways (25) and they are often coded in the same polycistronic unit, the operon (50). Many operons code for paths of metabolic networks, typically without skipped steps (122). Yet, the original paradigmatic *lac* locus in *E. coli* shows an even more interesting story. First, it is composed of two, not one, transcription units, where the regulator is transcribed apart but placed contiguously in the chromosome. The colocalization of an operon and its regulator is very frequent in bacterial genomes (40, 57). Second, the *lacZYA* operon contains a transporter and two enzymes, showing that functional neighborhood is not limited to connectivity in a metabolic network. Indeed, the most conserved operons code for proteins of similar functional classes even when they are not enzymes (24, 101). Third, the *lacZYA* operon is present in only one species, *E. coli*, among the first 500 completely sequenced genomes, and not even all strains of *E. coli* have the complete operon. In addition to being a celebrity and a paradigm, the *lac* operon is also a rarity.

Causes for the Existence of Operons

While wondering at the marvelous complexity of regulatory strategies, most researchers instinctively contemplate the regulatory model of operon evolution. The model sustains the proposition that functional neighbors are adaptively brought together in the chromosome for regulatory purposes. Given the historical role of operons in molecular genetics, it is perplexing

Cotranslational folding: concomitant folding of peptides at the moment of translation

that it took three decades for serious evolutionary questioning on the origin and maintenance of operons. In fact, the regulatory model raises at least three important questions: (*i*) Why are there operons when coregulation does not require them? (*ii*) How are genes brought together before coregulation has evolved? (*iii*) Why should neighbor operons frequently correspond to functional neighbors, as in the *lac* operon? In their landmark work, Lawrence & Roth proposed an alternative model wherein cluster formation and conservation is the result of selection on genes, not on organisms, to increase their fitness through gene transfer (60). Genes are massively transferred among most prokaryotic genomes and a cluster of genes performing neighborly functions has a much higher probability of successful transfer because it adds a functional module to a pre-existing structure. As a case in point, enzymes encoded in successfully laterally transferred operons tend to correspond to paths of metabolic pathways that are connected to the native ones (80). Under this model, operons are fitter because they allow seamless integration into the cellular networks of transferred genetic information. Although it has been coined the selfish operon model, in most situations the association is mutualistic, and the associated increase in cell fitness will effectively increase the frequency of the operon in the bacterial gene pool. The point of divergence between the regulatory and the selfish models is that the former emphasizes the advantage of cotranscription for regulatory purposes whereas the latter emphasizes the advantages of genome proximity for cotransfer of neighboring functions. Other models of operon evolution have been proposed but they have received far less attention, mainly because they do not fit available evidence (59).

Why Are There Operons?

Genes need not reside in operons to be successfully transferred or regulated in sophisticated ways. A theory to explain the creation and maintenance of operons must then explain why operons exist at all. Genes arising from different backgrounds are bound to have incompatible regulatory sequences. Their concatenation into an operon under the control of a single promoter is an all-or-nothing strategy: If the promoter works all genes will be expressed, if it does not then none will. When operons code for a single functional module or physical complex, as is often the case (24), only the expression of the whole has an adaptive value. But in most other cases, the advantage is less evident, and the dependence of all genes on a single promoter means that transfers lacking this single sequence will be unsuccessful for all the genes in the operon. In the absence of detailed modeling, the advantage of having operons under the selfish model is still open to debate. The regulatory model explains the existence of operons in several ways. First, the dependence of several genes on a single regulatory sequence puts this sequence under stronger selection and thus allows for the emergence of more complex regulatory strategies. Indeed, upstream regions of multigene transcriptional units have slightly more (~10%) regulatory signals than do regions upstream of single genes (87). Sharing a regulatory sequence also saves space and decreases the genetic load associated with selecting for a given motif. Second, in prokaryotes transcription-translation coupling is the rule and cotranscription allows colocated translation, which has been suggested to be adaptive, e.g., by allowing assembly of large complexes through cotranslational folding (109), or by allowing cell compartmentalization (21). Third, when different genes are to be expressed in exactly the same amount because they are part of a complex, transcription of all genes in a single transcript diminishes gene expression noise and ensures more precise stoichiometry. The most conserved operons code for proteins that interact physically (22), but in general the optimal expression level of each gene is not the same for all genes in an operon. This may be tackled by regulation at other levels, but adds complexity to the apparent simplicity of gene regulation by operons. Fourth, coincident mRNA degradation of genes in the same operon facilitates

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

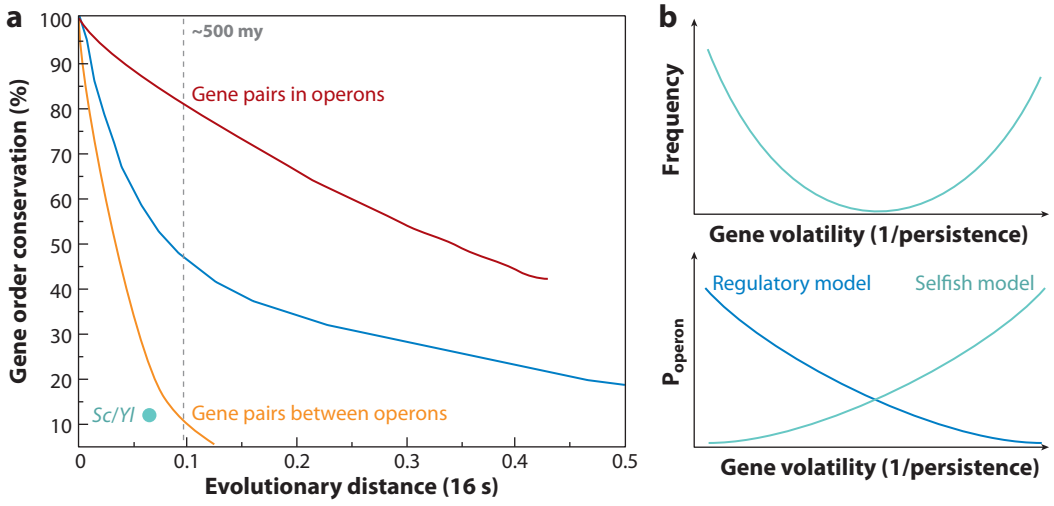


Figure 3

(*a*) Decrease in gene order conservation with divergence time for all genes (*blue line*), for pairs of genes in operons (*red*) and between operons (*orange*) [adapted from (93)]. The dashed line indicates a distance of ~500 million years (MY) and the green dot shows results for two yeasts (*Saccharomyces cerevisiae* and *Yarrowia lipolytica* (31)) that diverged ~300 Mya (108) (details in **supplementary material**; Follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**). These estimates are inaccurate and only give an order of magnitude of the time span involved. (*b*) Frequency of genes in a genome relative to volatility: most genes are either very persistent and present in most strains of a species or very volatile and present in few strains of a species (*top graph*). Schematic predictions of the selfish and regulatory models (*bottom graph*). The probability of a gene being in an operon is a function of gene volatility in both models but while in the selfish model the most volatile genes are more prone to be in operons, in the regulatory model the less volatile genes are more prone to be in operons (*bottom graph*). Volatility is the inverse of persistence, i.e., the probability that a gene of the pan genome is absent in a given genome of the species.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

the control of gene expression. The regulatory role of operons, although undisputed, does not necessarily signify that operons appear initially for regulatory purposes.

Where Are Operons Created?

The selfish theory claims that genes are brought together more frequently by rearrangements at the level of a mobile carrier of genetic information, e.g., an unstable plasmid, than at the level of the chromosomes (60). Whereas plasmids merge, split, and rearrange very frequently, bacterial chromosomes are highly stable, and significant shuffling of the core genome takes hundreds of millions of years (**Figure 3*a***). Hence, it might seem that it would take an excessively long time to bring two well-separated genes together simply by successive rearrangements. Yet genes can be brought together also by insertion and deletion of genetic material concomitant with a few large rearrangements and xenologous displacements. Available data show that most new operons containing native genes result from rearrangements and deletions of intervening genes, suggesting frequent operon formation in the chromosome (86). Even if operons are more likely to be formed in extrachromosomal elements, favoring the selfish operon model, this has to be weighted by the probability that genes actually meet in extrachromosomal elements and by the likelihood that the resulting operon is adaptive in the recipient genome. The available evidence shows that effectively selfish operons, e.g., poison-antidote systems or transposable elements, are rapidly lost (56, 115). On the other hand, operons created in situ may not make functional sense as a unit, and be subsequently split apart, but they have the appropriate regulatory

Supplemental Material

Core genome: a set of genes present in all strains of a species

Xenologous displacement: displacement of a native gene by a homologue acquired by horizontal transfer

Volatile genes: genes that once acquired by a genome are quickly lost, e.g., genes coding for very peripheral functions or selfish DNA

Pan genome: set of genes present in at least one genome of a species

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

sequences. Furthermore, each native gene has a higher probability of persisting in the chromosome because it is being individually selected for. Therefore, operons created by chromosomal events might be more frequently fixed, but further population genetics studies are required to assert this point.

Why Are There Conserved Contiguous Operons?

Genes in operons are under much stronger selection to remain together than genes that are contiguous in the chromosome but contained in different operons (24, 28). Yet, there is also evidence of weaker selection for contiguity between pairs of operons (93), some of which are so widely conserved that they have been called super- or uberoperons (58, 98). Although not cotranscribed, these structures are important elements of genome organization and may enlighten us on the evolutionary role of operons. The selfish operon theory readily explains why transferred genes need not be in one single operon to enjoy the selfish drive, the archetypical example being pathogenicity islands. But there may be regulatory advantages for contiguity between related operons. Pairs of divergently oriented operons show correlated expression levels and are more conserved than are convergent ones (57). This is because sometimes they share bidirectional regulatory regions that allow coregulation of the two operons (40, 117). The regulatory model might explain the existence of larger groups of operons, or pairs in other configurations, if closed nucleoid structure hinders gene expression. In this case, nearby positioning of related functions facilitates coexpression.

Key Predictions

The key prediction of the selfish theory is that the propensity of a gene to be in an operon should be positively correlated with its likelihood of being transferred (**Figure 3*b***). Therefore, laterally transferred genes should be particularly prone to be in operons and essential genes should stand alone (60). This prediction has been proven false, because the largest and most conserved operons concern essential genes such as those coding for ribosomal proteins (49, 79). Among the operons recently formed in the chromosome of *E. coli*, native genes are overrepresented and recently acquired genes most frequently form new operons with these (87). The selfish model predicts that sexually isolated genomes after a period of small population sizes and extensive rearrangements, such as *Buchnera*, should have smaller operons because they experience no horizontal transfer and have almost only kept essential functions. Instead, *Buchnera* have operons longer than those of the average bacterium (122), as expected if coregulation of essential functions were under strong selection. There are therefore many observations favoring the regulatory over the selfish model when it comes to key predictions. Yet, one must also consider the possibility that the selfish model might better explain the aggregating behavior of highly volatile genes, i.e., genes present in very few genomes within a clade (**Figure 3*b***). Even though these may constitute a small fraction of every genome, they still constitute a significant fraction of the pan genome and a major source of strain diversification. Essential genes stay for such long periods in the chromosome that they are bound to cluster at a given moment (30). Since these genes are rarely lost, the so-formed adaptive operons remain stable throughout long periods of time. At the other extreme, highly volatile genes correspond to functions that are often lost and rarely gained. They stay for such short periods in any given genome that there is little time to form adaptive operons in situ. The survival of these genes depends crucially on the success of horizontal transfer and thus the selfish model might better explain their presence in operons. Still to be quantified is the relative pertinence of each model to operons containing these poorly characterized genes. Ironically, although genomic data favor the regulatory model, one of its paradigms, the *lac* operon, has probably emerged in ways best explained by the selfish model.

GENOME ORGANIZATION BY NUCLEOID COMPACTION

The chromosome of *E. coli* is nearly 1000-fold compacted in a nucleoprotein complex called the nucleoid to fit around one fifth of the cell's volume (44). The cytosol has a high density of macromolecules and by the effect of excluded volume it contributes to nucleoid compaction (61). Macromolecular crowding is approximately constant throughout life cycles and growth conditions and does not involve sequence-specific interactions with the chromosome. As a result, it is relatively insensitive to chromosome organization. The two other determinants of nucleoid folding, negative supercoiling by topoisomerases and condensation by the attachment of nucleoid structure proteins, both shape and are shaped by chromosome organization. Negative supercoiling favors nucleoid condensation and is essential for the cell's survival, as it favors DNA unwinding and thus many cellular mechanisms interacting with DNA, most notably transcription. The nucleoid is highly condensed during rapid growth, when RNAP (RNA polymerase) concentrates in transcriptional foci, and much less so under starvation, when RNAP is distributed throughout the chromosome (36). There is thus an intimate association between genome organization and nucleoid structure via the distribution of highly expressed genes whose transcription affects supercoiling. The repertoire of DNA-binding structural proteins varies with growth rates and is associated with the topological remodeling of the nucleoid that is concomitant with changes in the distribution of RNAP (2). Although such proteins were thought not to recognize specific sequences, recent data show that at least H-NS binds well-defined DNA motifs (8). As a result, the distribution of motifs in the genome is expected to contribute to nucleoid structure, and, inversely, constraints on nucleoid structure will result in selection for biased distribution of motifs in the chromosome. The extent to which nucleoid structure affects and/or is passively affected by cellular processes has not been quantified. Yet, it has been suggested that nucleoid structure has a primordial role and that it leads gene order to adapt to the nucleoid folding, constituting a major barrier to genome change (16).

RNAP: RNA polymerase

H-NS: a small highly expressed chromatin-associated DNA binding protein

Nucleoid Structure

The nucleoid is structured into small supercoiled loops that are relaxed independently when DNA is interrupted. These domains of relaxation protect the chromosome from breaks that would otherwise lead to cell death by total loss of supercoiling. Experimental determinations of the average sizes of these domains differ between 10 kb (85) to 100 kb (119), with some studies indicating intermediate values (41, 99). The variance around these average values is very large, with the number of domains in the chromosome of *E. coli* estimated to vary between 12 and 400. These disparities have been attributed to inaccurate measurements (85), but they could also result from superstructures of domains that would react differently to different challenges. Small ~10-kb domains may be organized in higher-order structures if some barriers between domains are stronger than others or if there are sequence determinants for such superstructures. Gene distribution and orientation are consistent with a multiscale structure of the bacterial nucleoid imprinted in the form of genome organization (3). There is also experimental evidence that such superstructures exist. Based on the frequencies of intrachromosomal site-specific recombination, it has been proposed that the chromosome of *E. coli* is organized in four large macrodomains and two largely unstructured regions (112).

Interplay Between the Nucleoid and Genome Organization and Expression

The nucleoid is located in the center of the prokaryotic cell, RNAPs lie on its periphery, and ribosomes are found in the edges interacting with the inner membrane (63). The consequence of this arrangement is that DNA accessibility to the RNAP affects gene expression.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

Persistent genes: genes present in the majority of genomes of a clade, usually associated with housekeeping functions, e.g., essential genes

The average operon is less than 5 kb long, and transcription elongation of such an operon interferes with nucleoid structure at the level of the domain. Yet, for the RNAP to gain access to the chromosome, i.e., for transcription to start, higher levels of organization might also be implicated to allow DNA exposure to the RNAP: Once a region of the nucleoid is open for transcription, nearby operons may also be coexpressed.

The micro- and macrostructure of the nucleoid must be highly dynamic to tackle quick transitions between growth conditions and to allow transcription and replication. This coordination could be achieved if condensation is keyed by the *tabula rasa* effect of replication itself (85). From the point of view of genome organization, this has important consequences and raises a number of questions.

What Is the Effect of Nucleoid Structure on Gene Order via Gene Expression?

The expression of many genes is affected by changes in the level of supercoiling (82). Nucleoid proteins also influence gene expression because they have a preference for binding the AT-rich intergenic regions where transcription is regulated (35). This suggests recruitment of these proteins for regulatory purposes and leads to the association of nucleoid structure and gene expression. At a more global level, the folding of the nucleoid in domains puts into spatial contact distant chromosomal regions. If nucleoid structure influences transcription then long-range correlations may arise in gene expression patterns. Indeed, expression patterns correlate at short (<16 kb), medium (~100 kb), and long (~600–700 kb) distances in both *E. coli* and *Bacillus subtilis* (12, 51). These effects are small, which may result from a weak effect of the nucleoid periodicity on gene expression, from the weak periodicity of the nucleoid structure itself, or from noisy data. However, the similarity of patterns between such distantly related genomes strongly suggests a highly conserved underlying cause. The short-range correlations could result from the barriers between elementary domains of the nucleoid and/or from the effect of operons and superoperons. This could work in both senses, either the existence of operons and superoperons imposes domain structure via the remodeling effect of RNAP during cotranscription and/or nucleoid structure could be a driving force for the formation of superoperons. The longer-range correlations might result from higher scales of organization, e.g., from long-range interactions between regions of the chromosome that are linearly distant, but close together when the nucleoid is folded. This hypothesis requires a stable, reproducible, and evolutionarily meaningful macrodomain folding, as if it were directed by information present in the genome.

Is Nucleoid Structure Sequence Dependent?

The distances between regulators and their binding sites in *E. coli* show periodicities around 100 kb (54), compatible with the gene-expression periodicity mentioned above. The distribution of the most highly expressed and persistent genes, i.e., those whose presence is most highly conserved, also shows a periodicity in the chromosome of *E. coli* of ~100 kb (121). As these genes tend to be essential and associated with growth, they are often expressed under common genetic regulatory mechanisms, which could explain the consistency of the ~100-kb periodicities found for the different types of data. Although expression patterns can result from fortuitous folding of the chromosome in a given set of experiments, the periodicities between regulatory genes and their binding sites and among persistent genes are imprinted in the chromosome. It thus suggests selection for an arrangement of the nucleoid where periodically distributed genes could be coaccessed at the same time. Several models have been proposed that lead to such macroarrangements of the nucleoid (54, 118, 121), but it is unclear if they lead to structures resilient to the frequently large transfers and deletions of genetic information that will perturb chromosomal periodicity.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

Chromosomal rearrangements resulting in a mixture of different macrodomains have much more deleterious effects than inversions within them (29). This suggests the existence of selection underlying the discrimination between macrodomains. Preliminary data show that protein binding to specific DNA motifs is involved in the folding and individualization of the macrodomain surrounding the terminus of replication (F. Boccard, personal communication). Stable higher levels of nucleoid structure might then be under selection for adequate interactions between the chromosome and cellular processes, and this structure could partly be sequence dependent. If so, sequence evolution influences chromosome structure, but selection on nucleoid structure will also constrain chromosome dynamics and sequence evolution.

Might Nucleoid Structure Coevolve with Cellular Processes and the Environment?

Differences in the magnitude of nucleoid condensation among species affect replication and transcription patterns and may be adaptive for a variety of reasons. The chromosome of *Salmonella enterica* Typhimurium is more relaxed than that of *E. coli* (15), raising speculation that it might allow *Salmonella* to resist prophage induction or survive oxidative stress induced by macrophages (15, 55). Bacterial genome sizes vary between less than 200 kb to over 13 Mb, and cell sizes range from 0.2 to 750 μm. This inevitably entails very different degrees of nucleoid condensation. Compared to *E. coli*, the nucleoids of *Beggiatoa* have $\sim 10^6$ times more space to fold, which may or may not be used. On the other hand, some fast-growing *Mycoplasma* chromosomes must be at least about ten times more condensed to fit into their very small cells. Even if genomes are small and cells large, polyploidy may lead to high chromosome condensation. *Azotobacter* cells grow in size up to 10 times the volume of *E. coli* when they contain over 100 copies of the chromosome (71). The possibility of diverse and evolving levels of nucleoid structure should be borne in mind when analyzing the evolution of genome structure, as it will constrain gene expression and distribution.

REPLICATION AND ITS INTERACTIONS WITH GENE EXPRESSION AND SEGREGATION

The origin of replication is the only *cis*-acting essential region of the *E. coli* chromosome (53). Both replication forks start at the origin and replicate the chromosome, following opposite directions until they arrive at the terminus, the *dif* site, where decatenation or chromosome dimer resolution takes place (**Figure 1**). The chromosome is thus separated in two halves, the replichores, replicated by different forks. The passage of the forks remodels nucleoid structure and displaces all molecules physically interacting with the chromosome. While transcription shapes the chromosome structure at a local scale, replication, by its inherent asymmetries, does so at the scale of the replichores, i.e., of the entire genome.

Replication-Associated Gene Dosage Effects

The possibility of starting a new round of replication before the previous round finishes, i.e., of having simultaneous replication rounds, allows cells of *E. coli* to double every 20 min, whereas the chromosome takes three times longer to replicate. The estimated number of simultaneous replication rounds (R) is the ratio between the time required to replicate the chromosome and the time between two successive cell divisions. If R is close to zero, then chromosome replication rarely takes place in the cell. If R is 1, one replication round starts when the previous ends. When $R > 1$, cells experience multiple simultaneous replication rounds. A gene near the origin will be on average 2^R more abundant in the cell than a gene near the terminus (18). Naturally, R depends on the growth conditions, but while under optimal conditions some bacteria have low values of R, others can have $R > 3$. In the latter case, replication-associated gene

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

KOPS: FtsK orienting polar sequences

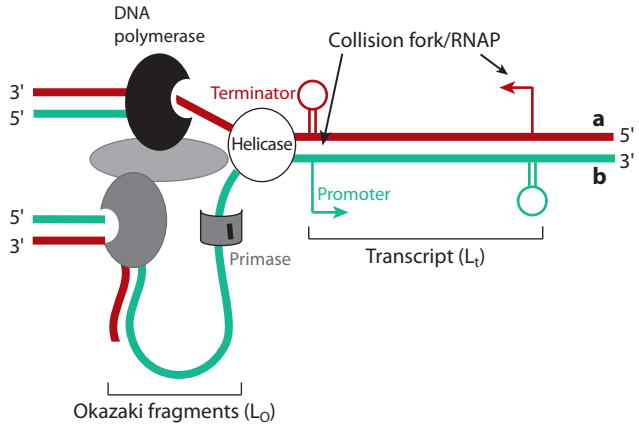
dosage is important as there are >3 simultaneous replication rounds in the cell and genes near the origin are >8 times more abundant in the cell than are genes near the terminus. This dosage effect is routinely used to map origins of replication in synchronized cultures, and in the absence of counteraction by genetic regulation it leads to higher expression of genes near the origin of replication (102, 105). Rearrangements changing the distance of a gene from the origin of replication will thus change its expression rate and affect optimal growth rates (11, 42). Although this is a purely mechanistic consequence of the process of replication in bacteria, it may be recruited for adaptive purposes. In particular, highly expressed genes near the origin will enjoy a replication-associated gene dosage effect allowing even higher expression levels. This gene dosage effect is more important in fast-growing bacteria, because R is higher and selection for quick growth more intense, and for the genes whose expression approaches saturation under exponential growth, such as RNAP, rDNA, and ribosomal proteins. These genes cluster systematically near the origin of replication in fast-growing bacteria (19). rDNA expression is regulated by the cellular concentration of free RNAP, and ribosomal protein expression is regulated by the cellular concentration of free rRNA. Strikingly, the relative positioning of these genes matches their regulatory dependences and in general RNAP genes are closer to the origin, followed by rDNA and then by ribosomal protein genes (19).

Replication, Segregation, and Gene Distribution

While replication and cell doubling are decoupled in bacteria, chromosome segregation is intimately associated with chromosome replication (76, 113). In *E. coli*, although some evidence points towards a period of cohesion between the newly formed chromosomes (1), other data show quick separation (76). In *Caulobacter*, chromosome segregation closely follows replication (113). Some authors have suggested the existence of a eukaryotic-like segregation apparatus in bacteria (4), but others have argued that genome organization might drive chromosome segregation (97). Because highly expressed genes accumulate near the origin of replication, this region becomes packed with RNAP and with ribosomes translating the nascent mRNA. The latter, because they constitute large complexes lying at the inner surface of the membrane, produce a powerful macromolecular exclusion effect, which may effectively pull the origins apart. Furthermore, rDNA, ribosomal proteins, and RNAP are coded in the leading strand (see next section). Therefore, RNAP is most frequently transcribing genes in a direction opposite to the origin of replication. Since RNAP is a potent molecular motor, this transcription bias could also result in quick separation of the origins (26). In both cases, gene expression allied to genome organization could contribute to chromosome segregation. Slow-growing bacteria lacking highly expressed genes near the origin of replication may not enjoy these effects. They may also not need them because in slow-growing bacteria the large lag time between replication rounds leaves ample time to segregate the chromosome by other means.

The decatenation and segregation of the newly replicated chromosomes to each of the daughter cells are highly accurate. Even the highly asymmetric segregation of sporulating *B. subtilis* chromosomes leaves less than 0.02% of anucleate cells (46). While the chromosome is replicated, cells elongate and a septum forms at the cell center. There, translocases, such as FtsK in *E. coli* and SpoIIIE in *B. subtilis*, directionally pump DNA into the daughter cells by recognizing motifs that point to the *dif* site. In *E. coli*, these motifs are called KOPS (FtsK orienting polar sequences) and their density is higher in the leading strand and increases toward the replication terminus, thereby indicating the direction of DNA translocation (7, 62). KOPS polarity constrains the chromosome dynamics near the terminus of replication because inversions lead to inversely polarized KOPS and therefore to a disruption of the

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.



Truncated mRNA (T):

$T_{lag} = 100\%$

$T_{lead} = (1\text{-}V_{RNAP}/V_{fork,\,lead}) \times 100 = 91\%$

RNAP exclusion time (t_x):

$t_{x,\,lag} = L_O/V_{fork,\,lag} + L_t/V_{fork,\,lag} + L_t/V_{RNAP} = 52.7$ s

$t_{x,\,lead} = L_t/V_{RNAP} - L_t/V_{fork,\,lead} = 36.5$ s

Effect on total expression (%):

$100 \times t_{x,\,lag}/t_d = 4.4\%$

$100 \times t_{x,\,lead}/t_d = 3.0\%$

Figure 4

Outcome of collisions between the fork and the RNA polymerase (RNAP) when genes are in the lagging (*red*) and leading (*aqua*) strands and collisions do not lead to replication arrests (in which case the impact of collisions is more important). Truncated mRNA refers to the fraction of aborted transcriptions while the fork passes in the transcribed region and assumes that all co-oriented collisions lead to transcription abortion (therefore the estimated difference is conservative). RNAP exclusion is the time when the region is unavailable for transcription. The effect on total expression of genes is the latter value time divided by the optimal doubling time (t_d). Computations and results (using parameters in *B. subtilis*) are detailed in **supplementary material**. (Follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**). The variables V_x refer to the rate of RNAP or the replication fork (nt/s).

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

decatenation process. As a result, some inversions in these regions are lethal, whereas deletions are viable, providing a striking example where genome organization prevails over gene content (88).

GENE STRAND BIAS AND THE ANTAGONISM BETWEEN REPLICATION AND GENE EXPRESSION

The replication fork synthesizes one DNA strand continuously, the leading strand, and the other semidiscontinuously, the lagging strand (**Figure 4**). The different replication mode of the two strands leads to different mutational patterns. As a result, the leading strand tends to be richer in G and the lagging strand richer in C in the vast majority of bacterial genomes (66), albeit from different mutational causes (89). This compositional bias allows the identification of the origin and terminus of replication and the delimitation *in silico* of bacterial replichores. Replicating strands differ not only in sequence composition but also in gene density. Most available evidence indicates that these two important discriminators between the leading and lagging strand are largely independent (74). While GC skews are caused by mutational biases, the overrepresentation of leading strand genes is created by natural selection on genome organization. I focus on the latter as GC skews have been extensively reviewed (32, 111).

The rates of nucleotide incorporation into macromolecules vary with growth conditions and between species. In *E. coli* the replication fork and the RNAP progress at 600–1000 nt/s and 30–80 nt/s, respectively (9). Since both polymerases are bound to the same template and replication and transcription occur simultaneously in dividing cells, collisions between them are inevitable. These collisions can be head on, if the transcribed gene is in the lagging strand, or co-oriented, if the gene is on the leading strand. The higher probability and harsher consequences of the former is thought to lead to gene strand bias, i.e., higher gene density in the leading strand. This bias in gene distribution was initially found among ribosomal genes and immediately associated with

Supplemental Material

GC skew: for a sequence with NG Guanines and NC Cytosines, GCskew = (NG–NC)/(NG+NC)

selection to avoid head-on collisions between the fork and RNAP transcribing these highly expressed genes (77). Head-on collisions slow the progression of the fork, and it was proposed that highly expressed genes, because they are more likely to be actively transcribed when the replication fork passes, would be particularly prone to be coded in the leading strand (10). In this model, selection against lagging strand genes is proportional to the number of collisions that transcription of these genes might generate, i.e., to their transcription rate in moments of replication, and to optimal growth rates, because replication slowdown is expected to be more deleterious in fast-growing bacteria. It was long accepted that gene strand bias was about the preference for highly expressed genes in the leading strand. This idea still echoes in some literature, but is incorrect: The essential genes, not the highly expressed genes, are highly over represented in the leading strand. In *B. subtilis*, ~95% of essential genes are on the leading strand independently of the level of expression, whereas 75% of the other genes are on the leading strand independently of their level of expression (94). These results hold for the other analyzed firmicutes and γ-proteobacteria, among which is *E. coli* (95), and raise two major questions.

Why Aren't There More Highly Expressed Genes in the Leading Strand?

In *E. coli*, collisions occur by direct physical contact between the fork and the RNAP and delay the fork much more if they are head on (33, 73). In *E. coli* plasmids, head-on collisions slow the fork especially if they take place at the promoter site, whereas co-oriented collisions slow the fork if they take place at the terminator site (72). In *B. subtilis*, inversion of gene strand bias retards replication by one third, but only in the presence of active transcription (116). Unexpectedly, the replication fork, while affected by transcription, replicates a highly expressed rDNA operon as quickly as the average gene (116). This suggests that in vivo the fork is not slowed by the sheer number of collisions, but simply by the existence of transcription, independently of the number of attached RNAP. One might speculate that this is in line with the observation that head-on collisions are much more deleterious than are co-oriented collisions at the promoter sites. If the difference between transcription in the leading and lagging strands is mostly due to the deleterious interactions of the RNAP with the fork at the promoter site, then the number of RNAP actively transcribing the gene is irrelevant. The only relevant parameter is the probability that the fork meets an RNAP-promoter interaction when it arrives at the regulatory region. Since most such interactions are abortive, they may occur all the time even for lowly expressed genes. In this scenario, selection for gene strand bias would affect all genes enduring frequent RNAP/promoter interactions at the time of replication and there would be no added selection pressure to code highly expressed genes in the leading strand. This fits the overall lack of strong overrepresentation of highly expressed genes in bacterial genomes.

If gene strand bias reflects selection against slower replication, then the bias should be high for fast-growing bacteria and low for the others. Genomes of firmicutes and mollicutes have much stronger gene strand bias than the genomes of other clades, showing close to 80% of leading strand genes. This might result from their peculiar replication and transcription machineries that could render them more fragile to head-on collisions (90). The other genomes show small strand biases among nonessential genes, which suggests that the effect of head-on collision on growth is very small, consistent with recent observations in *B. subtilis* (M. Itaya, personal communication) and *E. coli* (29), where only very large inversions of gene strand bias show important effects on cell growth. Furthermore, optimal doubling times are uncorrelated with gene strand bias (**Supplemental Figure 1**). The lack of strong preference for highly expressed genes on the leading strand, the low overall biases among nonessential genes, and the lack of association

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

Supplemental Material

between gene strand bias and optimal doubling times raise serious doubt about the link between gene strand bias and bacterial fitness by way of the effect of head-on collisions on the replication rate. Since replication and cell doubling are uncoupled in bacteria, a replication slowdown might not be deleterious if the interplay between replication and cell doubling compensates for that. A slower fork only results in lower growth if it implies a larger time lag between successive replication starts.

Why Are Essential Genes Preferably in the Leading Strand?

It has been suggested that if head-collisions lead more frequently to replication fork arrests, they may also result in higher local mutagenesis due to the fork salvage by homologous recombination (73). In this hypothesis, lagging strand essential genes would be avoided to limit the mutational load associated with head-on collisions. Less than 20% of replication rounds result in a replication arrest in *E. coli* (70), suggesting that collisions rarely, if ever, lead to replication arrests. Although highly expressed genes are very intolerant to sequence change, both synonymous and nonsynonymous essential genes are barely less tolerant to changes than the average gene (96). If local mutagenesis associated with head-on collisions were important, then it would lead to leading strand over representation of highly expressed genes and not of essential genes. One observes the inverse.

Instead of concentrating on the effects of collisions on replication and DNA, one might also contemplate the multiple effects of collisions on transcription and on mRNA. When the fork reaches a given coding region it has different effects depending on whether the transcript is on the leading or the lagging strand. Using the only complete homogenous experimental data set, that of *B. subtilis* (116), one can make a rough assessment of these effects (**Figure 4** and **Supplemental material**). These may be quite universal as the effects depend largely on the ratio between the rate of synthesis of RNA and DNA, which are ~0.05–0.1 in the fast-growing *B. subtilis* and *E. coli*, and ~0.2 in the slow-growing *Mycobacterium tuberculosis*, which has forks and RNAP 20 and 8 times slower, respectively (39, 43). As the fork passes, all lagging strand transcriptions are aborted and the region will be unavailable for transcription for some time. On the other hand, some leading strand transcripts will be finished before the fork displaces all RNAPs on an operon and the region will be unavailable for transcription for a shorter period of time. This has three possibly important consequences. First, the leading strand genes will have a slightly higher opportunity for being highly expressed. Yet, as shown above, there is no evidence for strong selection of highly expressed genes in the leading strand. Second, large aborted transcripts arise more frequently from lagging strand genes and can be translated into truncated peptides, which tend to produce negative dominants that can be highly toxic when involving essential functions (94). Third, collisions may increase gene expression stochastic noise, and this is more deleterious if genes are essential (75). The last factor will be particularly important if head-on collisions lead to more frequent replication fork arrests as this will render the genomic locus unavailable for transcription for a substantial period of time. Focusing on transcription abortion and DNA availability for transcription explains the lack of association between expression levels, growth rate, and gene strand bias, but at the cost of substantial speculation about the effects of truncated mRNA and gene expression noise. The data available on genome organization and on the effects of collisions between the replication fork and RNAP have produced an intriguing set of conflicting observations whose integration into a coherent theory begs for further experimental and evolutionary studies.

ORGANIZATION AND CHANGE

Bacterial genomes are highly fluid, yet remarkably stable. The rearrangement rate in *E. coli* is close to the genomic mutation rate ~10^{-3}–10^{-4} changes/(generation.genome) (42, 104).

Supplemental Material

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

When the genome of *E. coli* is compared to the one of *Salmonella enterica*, it shows near-saturation of synonymous positions and more than 10% changes in proteins. Yet, set aside insertions and deletions of genetic material, the two species genomes are almost colinear, showing that practically no rearrangement escaped purge by natural selection. Also, among 20 strains of *E. coli* ~18,000 different genes are found, although each *E. coli* has only around 4500 (E.P.C.R., unpublished data). Yet, most strains of *E. coli* have the exact same relative gene order among orthologues. Genomes mutate, change in size, and rearrange. Yet, large rearrangements are opposed by natural selection because they are particularly deleterious to genome organization. Somewhat surprisingly, they are more deleterious than many insertions and deletions of genes of the pan genome.

When selection for an organizational trait is weak, disruptive changes will be only weakly counterselected and some may become fixed. This is what happens for replication-associated gene dosage effects in slow-growing bacteria, which do not carry enough selective advantage to resist drift, and thus show little organization in this respect. In general, the observation of strong organizational features, as opposed to random distribution of genetic objects, is a good indication of selection for the mechanisms producing them.

Genome stability depends on not only selection for organizational traits, but also the overall efficiency of selection and chromosomal rearrangement rates. If effectively reproducing populations are small then selection is less efficient in purging slightly deleterious changes. In this case, the probability that mildly deleterious rearrangements get fixed is higher and genomes are less stable and less organized. Rearrangement rates vary between genomes because they depend on the extant recombination mechanisms and their targets, most notably DNA repeats. Therefore, there is a negative association between genome stability and repeat density (91), which is particularly striking if these repeats are transposable elements (103). The interplay between selection, its efficiency, and rearrangement rates can result in a variety of scenarios. *Buchnera* have low effective population sizes that could lead to instability via inefficient selection against chromosome rearrangements. Yet, because they lack repeats and homologous recombination, such rearrangements are expected to be extremely rare. As a result, the genomes are remarkably stable (107). When low population sizes are accompanied by the presence of recombinogenic elements, such as repeats or insertion sequences, then genomes are very unstable, e.g., as in some *Yersinia* and *Bordetella* (14, 81).

Genome stability can be analyzed experimentally or by comparative genomics. The two approaches give fundamentally different pieces of information. Experimental work allows rearrangement rates to be determined, while comparative genomics ascertain how rearrangements accumulated in the evolutionary history. Although the former is a better guide to understand events and mechanisms, the later explains better the effect of selection. Comparative genomics of rearrangement events has traditionally used either information about changes in local gene-order contexts or inferred the global changes arising in a lineage. The frequency with which pairs of colocalized genes in a genome have orthologues that are also colocalized in the other genome is a measure of the disruption of gene order at a local scale since the last common ancestor of the two genomes (45, 106) (**Figure 3*a***). Genomes that diverged recently are expected to share more extensive gene order than do distantly related ones. Therefore genome stability must be defined by calibrating the observed gene order conservation by the evolutionary time since the genomes diverged. In this sense, stability is the inverse of the observed rearrangement rate per unit of time. Stable genomes show higher gene order conservation after controlling for the effect of time, i.e., they show lower accumulation of rearrangements per unit of time (93). Quantification of genome stability allows testing of its hypothesized determinants such as pathogenicity (not significant), or repeat density (significant). It also shows that some clades are especially stable,

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

e.g., the *Buchnera*, whereas others are particularly unstable, e.g., the large clade of cyanobacteria (93). Many cyanobacteria have very large effective population sizes. Therefore, instability is certainly not caused by inefficient selection, but more likely is attributable to lack of selection for some organizational traits or very high rearrangement rates. This fits the observations that cyanobacteria have fewer and smaller operons (28) and frequently lack replication-associated organization (120). Why cyanobacteria should select weakly for these traits is unclear.

Selection for organizational traits leads to the preferential purge of some rearrangements over others. As a consequence, the trade-off between organization and change can sometimes be harmonized in cunning ways. These evolutionary strategies typically lead to the creation of regions of instability where most change takes place, while leaving the rest of the chromosome stable. Several genomes have such regions. In *Streptomyces* most essential, housekeeping and highly expressed genes are in the relatively stable center of the chromosome, whereas the genome becomes less and less stable toward the telomeres (17). In the unstable regions one can find over representation of repeats, transposable elements, and antibiotic production systems (5). Diversifying selection is thought to act upon the latter to circumvent natural acquisition of antibiotic resistance in nature. Therefore, selection for diversification of these elements implicates mechanisms that destabilize the chromosome and are confined to regions devoid of housekeeping functions and away from the origin of replication. An alternative to this strategy is found in genomes containing large amounts of plasmidic DNA. In *Borrelia*, the chromosome is remarkably stable, but the accompanying plasmids contain many repeated elements that generate important variability by recombination (13). In *Vibrio*, the smaller chromosome contains few essential genes but highly plastic sites. *Vibrio* species also count on superintegrons to fetch genes with minimal disruption of chromosome organization (100).

The manipulation of genomes allows quantification of the effects and relative frequencies of rearrangement events. Inversions that disrupt genes and operons are usually very deleterious, and most studies have controlled for these effects. Inversions leading to chromosome asymmetry, i.e., to replichores of different sizes, slow growth (42, 64), in direct proportionality with the asymmetry (M. Itaya, personal communication). In an asymmetric chromosome one replichore takes longer to replicate, leading to slower overall chromosome replication, and possibly posing problems for chromosome decatenation and segregation. Inversions that shift genes from the leading to the lagging strand can have very deleterious effects in *Lactococcus lactis* (11), which has almost 80% of genes in the leading strand, but only mildly deleterious effects in *E. coli*, which has 55% of genes in the leading strand (29). Rearrangements that change the polarity of KOPS elements near the terminus of replication are very deleterious because they complicate chromosome segregation and disrupt the macrodomain surrounding the terminus of replication (37, 68). Furthermore, some regions of the chromosome are much less accessible for recombination between them than others (34, 104), for example, because of the large macrodomains that serve as recombination insulators (112). Inversions that lead to the mixture of macrodomains are not only rare but also very deleterious (29). This limits the inference of selection for genome organization from the analyses of rearrangements, because it is difficult to distinguish low recombination frequencies from deleterious inversions, i.e., to distinguish mutational from selective effects. Two of the macrodomains in *E. coli* surround the origin and the terminus of replication, and inversions within these domains that are symmetric to the origin of replication are the most frequent type found in natural populations (27, 65, 110). The high frequency of these inversions also stems from their reduced negative effect on the large-scale organization of the chromosome. Indeed, these inversions do not disrupt any organizational feature associated with replication or

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

segregation and as long as they do not affect genes, operons, or superoperons they are close to neutral.

Prokaryotic genomes, despite showing intolerance toward inversions, are remarkably permissive to lateral transfer, because these events do not dramatically affect gene order or the large-scale organization of genomes. The elimination of recently acquired sequences from genomes can increase genome stability, because it leads to the removal of transposable elements, prophages, and other generic DNA repeats (84). But if the inserted DNA is inert, the effects of lateral transfer can be surprisingly neutral. The insertion of a nearly complete 3.5 Mb *Synechocystis* genome in scattered pieces inside the 4.2 Mb genome of *B. subtilis* had little phenotypic effect (48) as long as insertions did not disorganize the genome relative to replication. Similarly, multiple replicons can merge and split with little phenotypic effect (13). The chromosomes of *Sinorhizobium* spontaneously cointegrate with few noticeable growth changes (38). Wild-type *B. subtilis* has one single circular chromosome that could be artificially split into two autonomous replicons (47). Chromosomes can even be made circular if linear, as *Streptomyces* (114), or linear if circular, as *E. coli* (20). In the latter case, linearization was attempted at different points and only when both arms of the chromosome had equal length, with the origin of replication at the center, was growth indistinguishable from the wild type. This demonstrates the potential plasticity of the bacterial chromosome when organizational features are respected. Surprisingly, the linearized chromosome of *E. coli* is more robust to losses in the segregation apparatus (20). These results pose obvious questions. If such dramatic changes are possible, why are they rarely found in nature? Maybe subtle differences in fitness among variants exist but are hard to pinpoint experimentally. This may be why although *Sinorhizobium* replicons spontaneously cointegrate these variants are not usually found in nature, or why linear *E. coli* chromosomes have not been found so far. Another possibility is that because these genomic manipulations accounted precisely for genome organization, they are very unlikely to occur in nature by chance alone. Alternatively, we may have not sufficiently explored the natural variability of chromosome structure.

ARCHAEA

While the organization of the genome of archaea has received far less attention, it shares some similarities with bacteria. Notably, archaea also have conserved operons and transcription-translation coupling. Therefore, gene expression also organizes the chromosome. Fewer archaeal genomes have identifiable replication-associated genome organization. Some species have multiple synchronous origins of replication (69), whereas others have facultative origins (78), which so far have not been found among bacteria. Yet, when it is possible to distinguish between a leading and a lagging strand, the former was found to have more genes and be G-richer, as in bacteria (67). Reports also suggest that the presence of multiple similar chromosomes in a cell may be more frequent in archaea than in bacteria (83). Unfortunately, our current ignorance of many basic cellular mechanisms of archaea complicates the analyses of their effect on genome organization. Other points were deliberately overlooked in this review. A particularly studied element of genome organization concerns sequence composition (52). There are several updated articles on G+C content (6), strand asymmetries (111), and compositional heterogeneities (23).

CONCLUSION

Work over the past decade has exposed the puzzling evolutionary dynamics of prokaryotic genomes that allows them to be highly organized but also extremely plastic. This reflects the adaptability of prokaryotes to diverse and sometimes extreme environments, from hot spring to macrophages, with exceedingly fast replication rates, while using a wide diversity of coordinated metabolic pathways. Any cellular process interacting with the chromosome has been found to leave an imprint in it. If we

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

can elicit how organization comes about, we should also be able to draw inferences about cellular processes and ecotypes from the study of genome organization.

SUMMARY POINTS

1. Chromosomes are organized by their interactions with cellular processes.
2. While evidence suggests that operons and superoperons evolved mostly for regulatory purposes, the facility of cotransfer of neighboring functions may play a role in the evolutionary success of volatile genes.
3. The nucleoid is structured at different scales from 10 kb to 1 Mb. Some sequence determinants of nucleoid folding are being determined, suggesting coevolution of genome organization and nucleoid structure.
4. Gene expression and chromosome replication interact synergistically when leading to the selection of highly expressed genes near the origin of replication in fast-growing bacteria.
5. Gene expression and chromosome replication interact antagonistically when replication forks and RNA polymerase collide, resulting in selection of leading strand genes, especially among those coding for essential functions.
6. Comparative genomics, genomic manipulations, and synthetic biology are unraveling the mechanisms and effects of genome organization through the analysis of genome dynamics.
7. While bacterial genomes are immensely fluid in terms of gene repertoires, they are extremely conservative in terms of chromosome organization.

FUTURE ISSUES

1. Further testing of theories for the evolution of operons requires the use of population genomics approaches to inquire into the adaptive creation and dynamics of operons.
2. What is the interplay between operon and superoperon organization and nucleoid structure?
3. How are nucleoid domains structured into larger elements and how do these interfere with other cellular processes and with chromosome organization?
4. Is nucleoid folding driven by sequence? If so, which motifs and mechanisms are implicated?
5. Experimental work has demonstrated an important effect of head-on collisions on fork progression. Then why aren't highly expressed nonessential genes strongly overrepresented in the leading strand? And why are essential genes systematically in the leading strand, even when weakly expressed?
6. Given the complexity of factors shaping genome organization, mathematical models are needed to understand the effects of experimental genomic manipulations.
7. Why aren't there more linear chromosomes among bacteria?

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

8. Understanding how cellular processes shape the chromosome is a prerequisite for drawing inferences about cellular processes by studying chromosome organization.

DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Work in my lab benefits from a G5 grant from Institut Pasteur and from grant ANR-07-GMGE-004 from ANR. I'm greatly indebted to Antoine Danchin, Frederic Boccard, Philippe Glaser, Marie Touchon, and Todd Treangen for discussions. I further thank F. Boccard and Mitsuhiro Itaya for sharing unpublished results. I apologize to all colleagues whose work was not (sufficiently) cited due to space limitations.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

LITERATURE CITED

1. Adachi S, Fukushima T, Hiraga S. 2008. Dynamic events of sister chromosomes in the cell cycle of *Escherichia coli*. *Genes Cells* 13:181–97
2. Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* 181:6361–70
3. Audit B, Ouzounis CA. 2003. From genes to genomes: universal scale-invariant properties of microbial chromosome organisation. *J. Mol. Biol.* 332:617–33
4. Bates D, Kleckner N. 2005. Chromosome and replisome dynamics in *E. coli*: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell* 121:899–911
5. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–47
6. Bentley SD, Parkhill J. 2004. Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* 38:771–91
7. Bigot S, Saleh OA, Lesterlin C, Pages C, El Karoui M, et al. 2005. KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. *EMBO J.* 24:3770–80
8. Bouffartigues E, Buckle M, Badaut C, Travers A, Rimsky S. 2007. H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.* 14:441–48
9. Bremer H, Dennis PP. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ed. FC Neidhardt, et al., pp. 1553–69. Washington, DC: ASM Press. 2nd ed.
10. Brewer B. 1988. When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* 53:679–86
11. Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P, Le Bourgeois P. 2004. Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol. Microbiol.* 52:511–22
12. Carpentier AS, Torresani B, Grossmann A, Henaut A. 2005. Decoding the nucleoid organisation of *Bacillus subtilis* and *Escherichia coli* through gene expression data. *BMC Genomics* 6:84
13. Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, et al. 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
14. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, et al. 2004. Insights into the evolution of Yersinia pestis through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* 101:13826–31

15. Champion K, Higgins NP. 2007. Growth rate toxicity phenotypes and homeostatic supercoil control differentiate *Escherichia coli* from *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 189:5839–49
16. Charlebois RL, St Jean A. 1995. Supercoiling and map stability in the bacterial chromosome. *J. Mol. Evol.* 41:15–23
17. Choulet F, Aigle B, Gallois A, Mangenot S, Gerbaud C, et al. 2006. Evolution of the terminal regions of the streptomyces linear chromosome. *Mol. Biol. Evol.* 23:2361–69
18. Cooper S, Helmstetter CE. 1968. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* 31:519–40
19. Couturier E, Rocha EPC. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* 59:1506–18
20. Cui T, Moro-oka N, Ohsumi K, Kodama K, Ohshima T, et al. 2007. *Escherichia coli* with a linear genome. *EMBO Rep.* 8:181–87
21. Danchin A, Guerdoux-Jamet P, Moszer I, Nitschke P. 2000. Mapping the bacterial cell architecture into the chromosome. *Philos. Trans. R. Soc. London Ser. B* 355:179–90
22. Dandekar T, Snel B, Huynen M, Bork P. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23:324–28
23. Daubin V, Perriere G. 2003. G+C structuring along the genome: a common feature in prokaryotes. *Mol. Biol. Evol.* 20:471–83
24. de Daruvar A, Collado-Vides J, Valencia A. 2002. Analysis of the cellular functions of *Escherichia coli* operons and their conservation in *Bacillus subtilis*. *J. Mol. Evol.* 55:211–21
25. Demerec M, Hartman P. 1959. Complex loci in microorganisms. *Annu. Rev. Microbiol.* 13:377–406
26. Dworkin J, Losick R. 2002. Does RNA polymerase help drive chromosome segregation in bacteria? *Proc. Natl. Acad. Sci. USA* 99:14089–94
27. Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1:11.1-.9
28. Ermolaeva MD, White O, Salzberg SL. 2001. Prediction of operons in microbial genomes. *Nucleic Acids Res.* 29:1216–21
29. Esnault E, Valens M, Espeli O, Boccard F. 2007. Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genet.* 3:e226
30. Fang G, Rocha EP, Danchin A. 2008. Persistence drives gene clustering in bacterial genomes. *BMC Genomics* 9:4
31. Fischer G, Rocha EPC, Brunet F, Vergassola M, Dujon B. 2006. Highly variable rates of genome rearrangements between Hemiascomycetous yeast lineages. *PloS Genet.* 2:e32
32. Frank AC, Lobry JR. 1999. Asymmetric patterns: a review of possible underlying mutational or selective mechanisms. *Gene* 238:65–77
33. French S. 1992. Consequences of replication fork movement through transciption units in vivo. *Science* 258:1362–65
34. Garcia-Russell N, Harmon TG, Le TQ, Amaladas NH, Mathewson RD, Segall AM. 2004. Unequal access of chromosomal regions to each other in *Salmonella*: probing chromosome structure with phage lambda integrase-mediated long-range rearrangements. *Mol. Microbiol.* 52:329–44
35. Grainger DC, Hurd D, Goldberg MD, Busby SJ. 2006. Association of nucleoid proteins with coding and noncoding segments of the *Escherichia coli* genome. *Nucleic Acids Res.* 34:4642–52
36. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ. 2005. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc. Natl. Acad. Sci. USA* 102:17693–98
37. Guijo MI, Patte J, del Mar Campos M, Louarn JM, Rebollo JE. 2001. Localized remodeling of the *Escherichia coli* chromosome: the patchwork of segments refractory and tolerant to inversion near the replication terminus. *Genetics* 157:1413–23
38. Guo X, Flores M, Mavingui P, Fuentes SI, Hernandez G, et al. 2003. Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. *Genome Res.* 13:1810–17
39. Harshey RM, Ramakrishnan T. 1977. Rate of ribonucleic acid chain growth in *Mycobacterium tuberculosis* H37Rv. *J. Bacteriol.* 129:616–22

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

40. Hershberg R, Yeger-Lotem E, Margalit H. 2005. Chromosomal organization is shaped by the transcription regulatory network. *Trends Genet.* 21:138–42
41. Higgins NP, Yang X, Fu Q, Roth JR. 1996. Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J. Bacteriol.* 178:2825–35
42. Hill CW, Gray JA. 1988. Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* 119:771–78
43. Hiriyanna KT, Ramakrishnan T. 1986. Deoxyribonucleic acid replication time in *Mycobacterium tuberculosis* H37 Rv. *Arch. Microbiol.* 144:105–9
44. Holmes VF, Cozzarelli NR. 2000. Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc. Natl. Acad. Sci. USA* 97:1322–24
45. Huynen MA, Bork P. 1998. Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* 95:5849–56
46. Ireton K, Gunther NWt, Grossman AD. 1994. spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. *J. Bacteriol.* 176:5320–29
47. Itaya M, Tanaka T. 1997. Experimental surgery to create subgenomes of *Bacillus subtilis* 168. *Proc. Natl. Acad. Sci. USA* 94:5378–82
48. Itaya M, Tsuge K, Koizumi M, Fujita K. 2005. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl. Acad. Sci. USA* 102:15971–76
49. Itoh T, Takemoto K, Mori H, Gojobori T. 1999. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. Evol.* 16:332–46
50. Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318–56
51. Jeong KS, Ahn J, Khodursky AB. 2004. Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol.* 5:R86
52. Karlin S, Mrazek J, Campbell AM. 1997. Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* 179:3899–913
53. Kato J, Hashimoto M. 2007. Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3:132
54. Kepes F. 2004. Periodic transcriptional organization of the *E. coli* genome. *J. Mol. Biol.* 340:957–64
55. Khodursky AB. 2007. Evolution, adaptation, and supercoiling. *J. Bacteriol.* 189:5789–91
56. Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* 29:3742–56
57. Korbel JO, Jensen LJ, von Mering C, Bork P. 2004. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat. Biotechnol.* 22:911–17
58. Lathe WC, Snel B, Bork P. 2000. Gene context conservation of a higher order than operons. *Trends Biochem. Sci.* 25:474–79
59. Lawrence JG. 2003. Gene organization: selection, selfishness, and serendipity. *Annu. Rev. Microbiol.* 57:419–40
60. Lawrence JG, Roth JR. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–60
61. Lerman LS. 1971. A transition to a compact form of DNA in polymer solutions. *Proc. Natl. Acad. Sci. USA* 68:1886–90
62. Levy O, Ptacin JL, Pease PJ, Gore J, Eisen MB, et al. 2005. Identification of oligonucleotide sequences that direct the movement of the *Escherichia coli* FtsK translocase. *Proc. Natl. Acad. Sci. USA* 102:17618–23
63. Lewis PJ, Thaker SD, Errington J. 2000. Compartmentalization of transcription and translation in *Bacillus subtilis*. *EMBO J.* 19:710–18
64. Liu GR, Liu WQ, Johnston RN, Sanderson KE, Li SX, Liu SL. 2006. Genome plasticity and ori-ter rebalancing in *Salmonella typhi*. *Mol. Biol. Evol.* 23:365–71
65. Liu SL, Sanderson KE. 1996. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc. Natl. Acad. Sci. USA* 93:10303–8
66. Lobry JR. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* 13:660–65

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

67. Lopez P, Philippe H. 2001. Composition strand asymmetries in prokaryotic genomes: mutational bias and biased gene orientation. *C. R. Acad. Sci. III* 324:201–8
68. Louarn JM, Bouche JP, Legendre F, Louarn J, Patte J. 1985. Characterization and properties of very large inversions of the *E. coli* chromosome along the origin-to-terminus axis. *Mol. Gen. Genet.* 201:467–76
69. Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R. 2004. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. USA* 101:7046–51
70. Maisnier-Patin S, Nordstrom K, Dasgupta S. 2001. Replication arrests during a single round of replication of the *Escherichia coli* chromosome in the absence of DnaC activity. *Mol. Microbiol.* 42:1371–82
71. Maldonado R, Jimenez J, Casadesus J. 1994. Changes of ploidy during the *Azotobacter vinelandii* growth cycle. *J. Bacteriol.* 176:3911–19
72. Mirkin EV, Castro Roa D, Nudler E, Mirkin SM. 2006. Transcription regulatory elements are punctuation marks for DNA replication. *Proc. Natl. Acad. Sci. USA* 103:7276–81
73. Mirkin EV, Mirkin SM. 2005. Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell Biol.* 25:888–95
74. Necsulea A, Lobry JR. 2007. A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.* 24:2169–79
75. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441:840–46
76. Nielsen HJ, Youngren B, Hansen FG, Austin S. 2007. Dynamics of *Escherichia coli* chromosome segregation during multifork replication. *J. Bacteriol.* 189:8660–66
77. Nomura M, Morgan EA. 1977. Genetics of bacterial ribosomes. *Annu. Rev. Genet.* 11:297–347
78. Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T. 2007. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet.* 3:e77
79. Pal C, Hurst LD. 2004. Evidence against the selfish operon theory. *Trends Genet.* 20:232–34
80. Pal C, Papp B, Lercher MJ. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* 37:1372–75
81. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* 35:32–40
82. Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR. 2004. Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.* 5:R87
83. Poplawski A, Bernander R. 1997. Nucleoid structure and distribution in thermophilic Archaea. *J. Bacteriol.* 179:7625–30
84. Posfai G, Plunkett G 3rd, Feher T, Frisch D, Keil GM, et al. 2006. Emergent properties of reduced-genome *Escherichia coli*. *Science* 312:1044–46
85. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR. 2004. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* 18:1766–79
86. Price MN, Arkin AP, Alm EJ. 2006. The life-cycle of operons. *PLoS Genet.* 2:e96
87. Price MN, Huang KH, Arkin AP, Alm EJ. 2005. Operon formation is driven by coregulation and not by horizontal gene transfer. *Genome Res.* 15:809–19
88. Rebollo JE, François V, Louarn JM. 1988. Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Proc. Natl. Acad. Sci. USA* 85:9391–95
89. Rocha EP, Touchon M, Feil EJ. 2006. Similar compositional biases are caused by very different mutational effects. *Genome Res.* 16:1537–47
90. Rocha EPC. 2002. Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends Microbiol.* 10:393–96
91. Rocha EPC. 2003. DNA repeats lead to the accelerated loss of gene order in Bacteria. *Trends Genet.* 19:600–4
92. Deleted in proof
93. Rocha EPC. 2006. Inference and analysis of the relative stability of bacterial chromosomes. *Mol. Biol. Evol.* 23:513–22

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

94. Rocha EPC, Danchin A. 2003. Essentiality, not expressiveness, drives gene strand bias in bacteria. *Nat. Genet.* 34:377–78
95. Rocha EPC, Danchin A. 2003. Gene essentiality as a determinant of chromosomal organization in bacteria. *Nucleic Acids Res.* 31:6570–77
96. Rocha EPC, Danchin A. 2004. An analysis of determinants of protein substitution rates in bacteria. *Mol. Biol. Evol.* 21:108–16
97. Rocha EPC, Fralick J, Vediyappan G, Danchin A, Norris V. 2003. A strand-specific model for chromosome segregation in bacteria. *Mol. Microbiol.* 49:895–903
98. Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI, et al. 2002. Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res.* 30:2212–23
99. Romantsov T, Fishov I, Krichevsky O. 2007. Internal structure and dynamics of isolated *Escherichia coli* nucleoids assessed by fluorescence correlation spectroscopy. *Biophys. J.* 92:2875–84
100. Rowe-Magnus DA, Guerout AM, Biskri L, Bouige P, Mazel D. 2003. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res.* 13:428–42
101. Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J. 2000. Operons in *Escherichia coli*: genomic analyses and predictions. *Proc. Natl. Acad. Sci. USA* 97:6652–57
102. Schmid MB, Roth JR. 1987. Gene location affects expression level in *Salmonella typhimurium*. *J. Bacteriol.* 169:2872–75
103. Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M. 2000. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156:477–88
104. Segall A, Mahan MJ, Roth JR. 1988. Rearrangement of the bacterial chromosome: forbidden inversions. *Science* 241:1314–18
105. Sousa C, de Lorenzo V, Cebolla A. 1997. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* 143:2071–78
106. Tamames J. 2001. Evolution of gene order conservation in prokaryotes. *Genome Biol.* 2:0020.1.11
107. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, et al. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296:2376–79
108. Taylor J, Berbee M. 2006. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia* 98:838–49
109. Thanaraj TA, Argos P. 1996. Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* 5:1594–612
110. Tillier ER, Collins RA. 2000. Genome rearrangement by replication-directed translocation. *Nat. Genet.* 26:195–97
111. Touchon M, Rocha EP. 2007. From GC skews to wavelets: a gentle guide to the analysis of compositional asymmetries in genomic data. *Biochimie* 90:648–59
112. Valens M, Penaud S, Rossignol M, Cornet F, Boccard F. 2004. Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J.* 23:4330–41
113. Viollier PH, Thanbichler M, McGrath PT, West L, Meewan M, et al. 2004. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl. Acad. Sci. USA* 101:9257–62
114. Volff JN, Viell P, Altenbuchner J. 1997. Artificial circularization of the chromosome with concomitant deletion of its terminal inverted repeats enhances genetic instability and genome rearrangement in *Streptomyces lividans*. *Mol. Gen. Genet.* 253:753–60
115. Wagner A. 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. *Mol. Biol. Evol.* 23:723–33
116. Wang JD, Berkmen MB, Grossman AD. 2007. Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 104:5608–13
117. Warren PB, ten Wolde PR. 2004. Statistical analysis of the spatial distribution of operons in the transcriptional regulation network of *Escherichia coli*. *J. Mol. Biol.* 342:1379–90
118. Willenbrock H, Ussery DW. 2004. Chromatin architecture and gene expression in *Escherichia coli*. *Genome. Biol.* 5:252

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

119. Worcel A, Burgi E. 1972. On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* 71:127–47
120. Worning P, Jensen LJ, Hallin PF, Staerfeldt HH, Ussery DW. 2006. Origin of replication in circular prokaryotic chromosomes. *Environ. Microbiol.* 8:353–61
121. Wright MA, Kharchenko P, Church GM, Segre D. 2007. Chromosomal periodicity of evolutionarily conserved gene pairs. *Proc. Natl. Acad. Sci. USA* 104:10559–64
122. Zheng Y, Szustakowski JD, Fortnow L, Roberts RJ, Kasif S. 2002. Computational identification of operons in microbial genomes. *Genome Res.* 12:1221–30

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

Annual Review of Genetics

Volume 42, 2008

Contents

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.

Errata

An online log of corrections to *Annual Review of Genetics* articles may be found at http://genet.annualreviews.org/errata.shtml

Annu. Rev. Genet. 2008.42:211-233. Downloaded from www.annualreviews.org
Access provided by 176.111.79.153 on 04/08/20. For personal use only.