ISSN 0026-2617, Microbiology, 2007, Vol. 76, No. 4, pp. 383–392. © Pleiades Publishing, Ltd., 2007.
Original Russian Text © A.A. Prozorov, 2007, published in Mikrobiologiya, 2007, Vol. 76, No. 4, pp. 437–447.

REVIEW PAPERS

# Regularities of the Location of Genes Having Different Functions and of Some Other Nucleotide Sequences in the Bacterial Chromosome

**A. A. Prozorov**

*Vavilov Institute of General Genetics, Russian Academy of Sciences, ul. Gubkina 3, Moscow, 117809 Russia*
*E-mail: prozorov@vigg.ru*

Received September 21, 2006; in final form, December 7, 2006

**Abstrac**t—The review considers the results of genomic research performed over the last decade that shed light on the location in the bacterial chromosomes of genes having different functions. A tendency towards polarity of the chromosome composition is observed: vitally important genes tend to be concentrated in the region of replication origin (oriC), and their concentration decreases toward the region of replication termination (terC). An oppositely directed polarity (an increase near the terC region) is observed for the distribution of certain oligonucleotides involved in the process of chromosome recombination and segregation.

*Key words*: chromosome, replichores, replication, leading and lagging DNA strand, vitally important genes.

**DOI:** 10.1134/S0026261707040017

The last decade witnessed an outburst of studies on bacterial genomics. Since 1995, a great number of chromosomal genomes have been completely or partially sequenced, not to mention plasmids. The accumulation of voluminous experimental data in public databases allowed their subsequent computer processing to be carried out. It became possible to predict, with this or that degree of certainty, the function of the sequenced genes.

Due to these achievements, knowledge was gained about some general regularities of the structure and functioning of the bacterial chromosome (such regularities had been predicted previously, and some of them were confirmed). In particular, regularities (or tendencies) were revealed concerning the nonrandom distribution of genes with vitally important functions and of some other nucleotide sequences in relation to the regions of origin and termination of the replication of the circular bacterial chromosome. The present review is devoted to the analysis of the data available in this field. I have made an attempt to gather data concerning the chromosomes of eubacteria (the replication of archaeal chromosomes seems to obey to other rules, which are in part similar to those observed in eukaryotes).

The scope of this review does not allow me to consider certain important aspects to the extent they deserve. Along with the original works, I will discuss data contained in review papers, primarily the review by E. Rocha [1].

## GENERAL CONCEPTS OF THE STRUCTURE AND REPLICATION OF BACTERIAL CHROMOSOMES

Among different bacteria, chromosomes vary in size, structure, and even in their number per cell. A typical situation is the presence of a single chromosome, which corresponds to the whole genome of the cell (extrachromosomal elements are not taken into account). However, certain species (e.g., *Brucella* spp.) contain two chromosomes of different sizes per cell, and a strain of *Pseudomonas cepacia* has been found to contain three chromosomes per cell (3.4, 2.5, and 0.9 Mb). A trait distinguishing small chromosomes from large megaplasmids, which are comparable in size, is the presence in chromosomes of certain genes that are vitally important for the cell, e.g., of rRNA genes; the cell can be cured of the plasmid, but it cannot be cured of any of its chromosomes (see the reviews [2, 3]). It should be noted that the presence of two or three chromosomes in a bacterial cell cannot be considered polyploidy, since these chromosomes do not double each other and mostly contain different genes. The size of the chromosomal genome also varies; a typical size is about 2–4 Mb (4.65 Mb in *E. coli*; 4.12 Mb in *B. subtilis* 168; 1.83 Mb in *H. influentse* Rd, see review [4]). However, in some highly specialized obligately parasitic species, the chromosomal genome is much smaller (e.g., 0.58 Mb in *Mycoplasma genitalium*, urethritis causative agent). On the contrary, in some actinomycetes, which are free-living bacteria that form aerial

mycelium during growth on solid media, the genome is as large as 8 Mb (*Streptomyces lividans*). The largest bacterial genome among those sequenced is that of *Myxococcus xanthus* (9.2 Mb); it is considerably larger than the genome of the eukaryotic microsporidium *Spragula lophii* (6.2 Mb) (see review [5]).

The typical geometry of the bacterial chromosome is a true (covalently closed) circle. However, the spirochetes of the genus *Borrelia* (e.g., *Borrelia burgdorferi*, the causative agent of a variety of tick spirochetosis) have a linear chromosome, both DNA strands of which exhibit hairpin structures at their termini. In many actinomycetes, the chromosome is a pseudocircle that is closed not covalently, but at the expense of the interaction of protein molecules located at the termini of the linear DNA molecule.

Although over the last years the number of bacterial species whose genomes have been completely sequenced reached several hundred, comprehensive functional genomics data are available for only a limited number of species, including the classical models of bacterial genetics: *E. coli* and *B. subtilis.* These bacteria are typical in terms of the above-discussed parameters: they have a single circular chromosome of an intermediate size. Therefore, considering below the schemes of structure and replication of the bacterial chromosome and some regularities of the location of different genes in it, I will primarily proceed from data obtained for these two bacterial species (mentioning the most significant distinctions revealed in other eubacteria).

In most eubacteria, replication of the chromosome begins in the so-called oriC region and proceeds, more or less symmetrically, along the right and left semicircles (*replichores*). An increase to a critical level in the content of protein DnaA (product of *dnaA* gene, located in this region [6, 7]; see also review [8]) serves as a signal to the beginning of replication. The oriC region includes sites with so-called DNA boxes (usually about 9 bp in size) alternating with AT-rich 12- to 13-bp sequences. The oriC region of *B. subtilis* contains 15 DNA boxes, that of *E. coli* contains five, and some other bacteria exhibit up to 19 DNA boxes in this region [7]. Among them, certain genes occur, including *dnaA*. The oriC structure proved to be conserved among many bacteria. The DNA boxes are not coding regions, but they bind protein DnaA when its level reaches a threshold concentration. Binding of this protein is a signal for the initiation of replication, which occurs via unwinding of double-stranded DNA by DNA helicase. At this site, replication forks are formed, which then move along the right and left replichores to meet at the site of termination of replication, terC. This region is not strictly opposite oriC, at least in *E. coli* and *B. subtilis*: in both of these bacteria, the left replichore is somewhat longer than the right one.

In each of the replication forks, a replisome occurs, which is a complex of enzymes performing replication; the replisome consists of DNA polymerase III, DNA helicase, RNA primase, and ligase [7]. DNA polymerase is a very large enzyme with a complex hierarchical structure. The total complex is sometimes termed *holoenzyme*; in the holoenzyme, two core subenzymes are distinguished, each consisting of several subunits. It is known that DNA replication occurs in different ways on the two strands. On the leading strand, it occurs continuously in the 5'–3' direction, and on the lagging strand, DNA synthesis occurs in the opposite direction and with the formation of separate fragments (*Okazaki fragments*, which are several thousand nucleotides long in bacteria). The beginning of the synthesis of each fragment requires a newly formed short RNA strand, produced by RNA primase. Okazaki fragments are then ligated by DNA ligase. In *E. coli*, both core subenzymes can work on both the leading and lagging DNA strands [9]; in *B. subtilis* and some other bacteria, the core subenzymes differ in the structure of subunits and specialization: one of the subenzymes works on the leading strand, and the other on the lagging one ([9]; see also [11]).

Due to the peculiarities of the replication of DNA strands, contradictions may occur between the operation of DNA polymerase III and another important enzyme, RNA polymerase [11, 12]. While transcribing operons, RNA polymerase moves along DNA 20-fold more slowly than DNA polymerase III; thus, the latter enzyme constantly passes operons that are being transcribed. RNA polymerase moves along the leading DNA strand in the 3'–5' direction, which coincides with the direction of the replication of this strand; therefore, the transcription process is not affected. The situation is different when RNA polymerase transcribes the lagging strand, whose replication occurs in an opposite direction: a head-on collision between the two enzymes may result in the appearance of incomplete transcripts and even impairment of replication. This is probably the reason for the location of the coding strand of vitally important genes (e.g., rRNA genes) in the leading DNA strand [13] (also see below).

In an "ideal" case, a single replication round occurs, which culminates when the two replication forks come in touch in the terC region. There exist special mechanisms that prevent the start of a new round until the previous one has been completed. In *E. coli*, one such mechanism is hemimethylation: over one-third of the replication round, only old (template) DNA is methylated, and the newly synthesized DNA is not, and it cannot, therefore, serve as a template in a new round [14].

The process that prevents the start of the replication round out of order is termed *sequestration.* In *E. coli*, protein SeqA is responsible for it. In *seqA* mutants, the oriC region is many times doubled in the process of replication and has a multiforked structure [15]. In "normal" bacterial cells, *B. subtilis* in particular, the mechanism that prevents a premature replication round in often impaired; in this case, two new forks or even a

cluster of forks appear in the oriC region, resulting in an increase in the dose of the corresponding genes. This dose doubles even in bacteria with an unimpaired sequestration; in cells in which sequestration is impaired, it increases by a factor greater than two, and this increase is greater the closer the genes are to oriC. In *E. coli*, the dose of genes (particularly, of gene *purA*, which is immediately adjacent to the *dnaA* gene) was initially determined by comparing the transforming activity of proximal and distal genes, which reflected their content in the total DNA preparation. In this way, as early as in 1963, the first strong evidence of the polarity of replication in this bacterium was obtained [16]. The dose of the protein products corresponds to the gene dose (see review [7]). Thus, in *B. subtilis*, the level of the enzymes tryptophan synthetase, D-serine deaminase, tryptophanase, thymidine phosphorylase, and β-galactosidase showed good correlation with the proximity of these genes to oriC [18]. For *S. typhimurium*, a number of translocations of the *hisD* gene within the operon were obtained, and the study of these translocants revealed a clear-cut regularity: the greater the distance from *hisD* to oriC, the lower the level of the enzyme it encodes, histidinol dehydrogenase ([19], see Fig. 1). Several analogous research papers have been published (see review [17]). Below, it will be shown that an increase in the gene dose determined by gene location promotes accumulation of the most important RNAs and proteins, and thus seems to play a role in the regulation of cell metabolism.

terC is a rather large part of the chromosome (in *E. coli*, it occupies about 280 kb in each replichore, i.e., about 12% of the replichore length). The boundaries of oriC are marked with several 22-bp noncoding sequences located rather far away from the sites where the two replichores come into contact (see review [20]). In *B. subtilis*, three such sequences are located closer to the terminus of each of the replichores. After protein TUS (in *E. coli*) or protein RTR (in *B. subtilis*) [21] binds to Ter-sequences, they prevent penetration of the replication fork of the opposite replichore, providing for a kind of one-way traffic. In the center of terC, there is a 28-bp dif (deletion-induced filamentation) site. For a certain period of time, the daughter chromosomes are an interwoven catenated dimeric structure. In the sites adjacent to the dif site (the so-called dif-zone, occupying about 8 kb in each replichore), decatenation of this structure occurs, which precedes segregation (separation) of the daughter chromosomes. In this process, a complex of enzymes is involved, which includes resolvases XerC and XerD, topoisomerase IV, DNA gyrase, enzymes of homologous recombination, and one of the domains of protein Fsk [22–24]. This complex is functionally related to particular DNA sites, Chi and Rag (their location in the chromosome is considered below). Both of these sequences are most probably involved not in replication per se, but in postreplication events, namely, in the recombinational repair of impairments occurring during decatenation, etc. The decatenation process seems to be inevitably connected with constant DNA rearrangements and formation of deletions. The site name dif was given because deletions in this site, which occur rather frequently, may lead to the emergence of filamentous cells with impaired nucleoid distribution. In the next section, attention will be given to the absence of vitally important genes in the regions adjacent to dif.

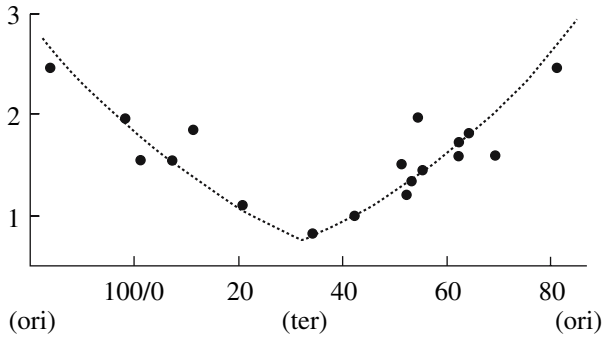


**Fig 1.** Activity of histidinol dehydrogenase (hisD gene product) in cell extract as dependent on the distance from this gene to the terC region in *Salmonella typhimurium* translocants. The ordinate axis shows activity in arbitrary units; the abscissa axis shows location on either of the terC sides and distances from terC in the genetic map in arbitrary units (modified from [19]).

In the experiments where phage λ was inserted in different sites of the *E. coli* chromosome and the frequency of its excision was then determined, it was shown that in the chromosome there exists a 4-kb site (mainly coinciding with the dif-active zone) where the excision frequency is a thousand times higher. This site was named TRZ (terminus recombination zone). Mutation in the *recA* gene abolished the increased excision frequency; thus, the TRZ site is a region of homologous hyperrecombination [25]. Most probably, decatenation, which is indispensable for the separation of replicated but still interwoven chromosomes, provides especially favorable conditions for recombination and/or itself occurs at the expense of recombination. In his review [26], T. Hill points out that it is difficult to understand what advantages a cell can gain from the presence of a permanently unstable genome region, and suggests that the hyperrecombination in the terminal region may be just the cost that bacteria pay for the replication of their circular chromosome.

In bacteria with linear chromosomes, if there are terminal hairpin-like structures at their termini, an oriC-like site is located in the middle of the chromosome. Both replication waves start at this site, proceed in opposite directions, and culminate at the terminal structures, which, in this case, play the terC role. In bacteria with pseudocircular chromosomes (e.g., actinomycetes), which are closed not covalently but by means of connecting proteins, two types of replication seem to be possible: from the center, where an oriC-like region is located, to the termini; and from the termini to the

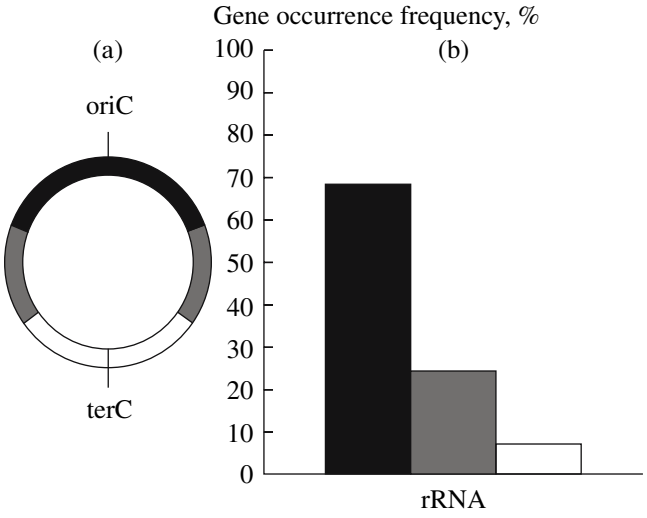


**Fig. 2.** Averaged occurrence frequency of ribosomal genes in different chromosomal regions in 68 bacterial species. (a) The chromosome is conventionally subdivided into equal-sized regions: black color shows the region adjacent to oriC and white color the region adjacent to terC; gray color shows the intermediate region. (b) Columns show % of ribosomal genes occurring in particular regions (modified from [11]).

center (for details, see the reviews [2, 4, 5] and references therein).

## LOCATION OF GENES OF DIFFERENT IMPORTANCE RELATIVE TO THE CHROMOSOMAL ORIC AND TERC REGIONS

From the previous section, it can be seen that the bacterial chromosome replication mechanism determines the existence of "favorable" and "unfavorable" gene locations. The most important factors are the distance from oriC, which influences the order in which the gene (and, consequently, its product) is duplicated, and the location of the transcribed sequence on the leading or lagging DNA strand. A natural hypothesis is that the genes most important for the cell should be located so as to provide optimal conditions for their functioning. In George Orwell's *Animal Farm*, the commandment was declared that "All animals are equal, but some animals are more equal than others." But what is the criterion to determine which genes are "more equal than others"?

The genes that undoubtedly rank first among the candidates are those that determine the ribosome structure. The ribosome is a vitally important cellular organelle, and uninterrupted synthesis of its components, rRNA first of all, is vitally important for the bacterial cell. Indeed, in the period of *E. coli* transition to exponential growth, ribosomal RNAs account for more than half of all the transcriptional activity of the cells, although the corresponding operons constitute only 0.5% of the chromosomal genome [27, 28]. As early as in 1981, it was noted that rRNA operons in enterobacteria are mainly located near oriC ([29]; see also the references in the reviews [30–33]). In *E. coli*, there are seven operons coding for rRNA; five of them are located near oriC (in both replichores), and two occur near the middles of the replichores (rrnC, A, B, E are near oriC in the right replichore, rrnD is near oriC in the left replichore, and rrnH and rrnG are in the middles of the right and left replichores). All these operons exhibit similar organization: two tandem repetitions, 16S rRNA gene, tRNA spacer, 23S rRNA gene, and 5S rRNA gene [27, 34]. The chromosome of *B. subtilis* contains ten rRNA operons [35]. The operons D, A, J, W, I, H, and G are very close to oriC (D and A are almost immediately adjacent to it). Cytological studies with visualization of gene products in *B. subtilis* showed that the synthesis of the rRNA O operon occurs near the poles of the dividing cell, in the region of gene *spoDJ* location and the immediate vicinity of oriC. The fluorescent spots of the products of operon O and the *spoDJ* gene are nearly merged [36, 37]; paraphrasing the postulate of dialectical materialism, it can now be said that the neighborhood of *spoDJ* and rrnO is the objective reality available to us as perception of light. An averaged distribution pattern of rRNA operons, obtained by analysis of 68 complete genome sequences available in public databases, turned out to correspond to the tendency already known in *E. coli* and *B. subtilis*: 70% of ribosomal operons were in the first third of both replichores, 22% were in their middle parts, and only 8% were closer to terC ([1]; see Fig. 2]).

A certain correlation exists between the growth rates of particular bacteria and the number of ribosomal operons contained in their genomes. Thus, *Vibrio natriegens* (a bacterium from the salty sludge of swampy seashore sites), which has a generation time as short as 10 min, has 13 ribosomal operons in the genome. Some rapidly growing pathogenic bacteria, such as *Vibrio cholerae, Vibrio parahaemolyticus*, the gas gangrene causative agent *Clostridium perfringens*, have eight, nine, and ten rRNA operons, respectively [38]. Unfortunately, I am unaware of any data on the location of ribosomal operons in the chromosomes of these bacteria. At the same time, in *Mycobacterium tuberculosis* strain H37RV, which has a low growth rate (generation time of 24 h), complete genome sequencing revealed only one rRNA operon. Notably, it is located closer to terC than to oriC [39].

The genes determining the synthesis of the numerous ribosomal proteins are usually located near oriC as well. Thus, in *B. subtilis*, 40 of 52 genes encoding ribosomal proteins are located not far from oriC; 38 of them occur in the right replichore, forming a large cluster only 120–140 kb away from oriC (each replichore of *B. subtilis* is 2100 kb in size). Two other genes are in the left replichore, very close to oriC [40].

What genes, apart from ribosomal genes, are vitally important for bacteria? I could a priori name a number of them (e.g., genes responsible for DNA replication), but this list would be to a certain extent subjective. An objective approach has been demonstrated in the work with *B. subtilis* 168, fulfilled by a large research team (99 authors!) [41]. This work used the database and mutant collection that were generated during complete sequencing of the genome of this bacterium [40]. Many researchers were involved in both of these projects. The following approach was the major one. Every gene of *B. subtilis* was, if this proved possible, inactivated at the expense of insertion mutagenesis, which was accomplished by addressed insertion into the chromosome of specially constructed plasmids of the pMITIN series [42]. These plasmids, which could not replicate in *B. subtilis* cells, carried, in particular, a selective marker of erythromycin resistance, the *lacZ* reporter gene, and an inducible promoter of this gene. Fragments of preamplified chromosomal genes meant to be inactivated were inserted in these plasmids, and the plasmids were used to transform competent *B. subtilis* cells (subsequent selection of transformants was based on erythromycin resistance). The insertion of a plasmid in the gene of interest occurred due to the presence in the plasmid of a homologous fragment; it occurred by a single crossing-over event and caused impairment of the gene structure. After the contact with plasmid DNA, competent cells were inoculated into LB medium, which is a rich medium that could provide for the growth of any auxotrophic mutant with impaired chains of metabolic reactions. However, mutations in fundamentally important genes should have led to lack of growth even in LB medium. Thus, the transformants in which the plasmid had been inserted in a vitally important gene could not grow on this medium. Additional verification of the importance of the genes was undertaken: since many genes are transcribed as operon members, an insertion of a plasmid in a near-promoter gene could inactivate the whole operon. Thus, the vitally important gene could be not the one inactivated, but a gene located further in the operon. To exclude this possibility, the plasmid was supplemented with its own promoter induced by isopropyl-β-thiogalactopyranoside (IPTG). The addition of IPTG to the medium restored the functioning of the bacterial operon, which was now transcribed from the plasmid promoter, and such transformants could grow.

Consideration of this scheme shows that certain reservations are necessary. Some genes may be present in the chromosome in several copies and in different sites (e.g., the above-discussed ribosomal operons), and the insertional inactivation of one of the copies may not be accompanied by inactivation of others. Then, many genes having important functions were not shown to be vitally important by the use of the scheme discussed. These are, e.g., *rec* genes; viable rec$^-$ mutants are long known, although no one doubts that the recombination ability is not a neutral trait. The same can be said about the sporulation and spore germination (*spo* and *ger*) genes, a large number of which are present in the *B. subtilis* chromosome. Mutations in most of these genes are not lethal; however, such mutants are not likely to survive in the natural habitat of this bacterium (soils of various natural landscapes). The evaluation of gene importance is also complicated by gene's polyfunctionality. For example, the already mentioned *spo* genes of *B. subtilis* are occasionally involved in processes other than sporulation, e.g., competence development, and it is difficult to say which of the functions is more important.

Nonetheless, the above-discussed work [41] produced a list containing 271 vitally important genes, or 6.6% of the total number of *B. subtilis* genes; hereafter, I will refer to this list as "the Kobayashi list" (according to the name of the first author of the publication). In this list, the authors included 132 genes revealed in that same work [41], 19 genes revealed by them in a similar way in earlier studies, 41 genes found in other investigations of *B. subtilis*, and 79 genes whose vital importance was not tested directly but was predicted based on their homology with vitally important genes of other bacteria (e.g., genes of ribosomal proteins). Beyond this list, there remained 3830 genes, including a huge number of genes of amino acid metabolism and almost all genes of sporulation and spore germination. The list did not include rRNA genes, which are evidently vitally important, because they could not be tested due to their presence in the chromosomes in several copies.

The genes from the Kobayashi list were subdivided into several categories according to their functions. I made an attempt to determine the location of these genes in the chromosome relative to oriC, specifically, to determine which of them are located in first third of the right and left replichores. The whole chromosome of *B. subtilis* is 4214 kb in size, and the first third of each replichore (60° of the 360° of the complete circle) corresponds to 720 kb. It should be said that this investigation was possible only for the genes (with a proved or predicted function) that have been depicted on the chromosome map of *B. subtilis* in [40].

The Kobayashi list begins with 27 genes involved in DNA metabolism (the main replication machinery; chromosome packing and segregation during cell division; methylation). 11 of these 27 genes are located in the first third of the replichores; in the oriC region itself (0° of the map), there are *dnaA* (initiation of replication), *dnaN*, and *dnaX* (beta, gamma, and theta subunits of DNA polymerase III) and *gyrA* and *gyrB* (A and B subunits of DNA gyrase). Somewhat farther from oriC are *holB* (delta subunit of DNA polymerase III), *dnaC* (replicational DNA helicase), the gene of the protein binding to single-stranded DNA, two genes of DNA methylation, and the gene of DNA ligase. Further in the list come 14 genes of RNA metabolism. Among them, close to oriC (11–12°), are RNA polymerase genes *rpo A*, *B*, *C* (alpha, beta, and beta-prime subunits of this

enzyme). In the very beginning of the left replichore is the *rapA* gene (encoding one of the components of ribonuclease P). In the voluminous category of vitally important genes of protein synthesis (95 genes), 52 are involved in the synthesis of ribosomal proteins; as already mentioned, 40 of them are in the first third of the chromosome. Among other genes of protein synthesis, in this region are 5 (out of 23) genes of tRNA synthetases, 6 genes of transcription factors, one of the genes regulating the folding of protein molecules, and one gene involved in protein secretion. Among the 42 genes involved in the synthesis of the cytoplasmic membrane and cell wall, 5 are within the first third of the chromosome. The other 70 genes in the Kobayashi list are related to glycolysis, respiration, nucleotide synthesis, synthesis of cofactors, etc. Of the 10 genes of nucleotide synthesis, 4 are in the first third of the chromosome; among 8 glycolysis genes, 5 occur there (although 4 are located at the very boundary of the first third of the left replichore). Finally, the *tagO* gene should be mentioned, which is involved in the synthesis of teichoic acid, a cell wall component, and is located in the beginning of the left replichore.

Thus, the location of a number of vitally important genes (genes DNA and RNA metabolism, genes of ribosomal protein synthesis, genes of nucleotide synthesis, and, possibly, glycolysis genes) exhibits the same tendency as the location of rRNA genes: they tend to be located in the beginning of replichores. This tendency is particularly pronounced for genes whose products are components of the replisome, genes encoding RNA polymerase subunits, and, of course, genes responsible for synthesis of ribosomal proteins. It should be noted that at least some of the DNA metabolism genes of other bacteria—*E. coli*, *P. putida*, *M. tuberculosis*, *Sinechocistis*—and even of *B. burgdorferi*, a spirochete with a linear chromosome, exhibit the same tendency in their location, although in different genomes the genes may be rearranged with respect to their order in the chromosome [5, 39, 43, 44].

And what about the genes that were not included in the Kobayashi list but are also important for the survival of bacteria (e.g., *rec* genes)? In the annotation of the *B. subtilis* genome, 5 *rec* genes are recognized, *recA*, *N*, *Q*, *R*, *F* [40] (initially, their number was much greater, but, due to the polyfunctionality of many of them, some were later assigned to other categories). The *recF* and *recN* genes (their "official" names in the annotation are "genes of DNA repair and genetic recombination" [40]) are quite close to oriC (*recF* is only 3 kb away from it). *recR* (having the same "official" name) is in the middle of the left replichore; *recQ* (gene of ATP-dependent DNA helicase) is close to *recR*. The famous *recA* gene, whose homologs have been found in virtually all pro- and eukaryotes (according to the annotation, it is the gene of a multifunctional protein involved in homologous recombination and DNA repair) unexpectedly occurs in the last third of the right replichore.

Let us now consider the cases of location of the Kobayashi-list genes in other regions of the *B. subtilis* replichores. If the numbers of vitally important genes in the first and last thirds of the chromosome are compared, the difference is not so great. As pointed out above, 11 vitally important genes of DNA metabolism occur in the first third of the chromosome; in the last third, 7 such genes occur. However, if the compared regions are narrowed to 5° (about 60 kb), the situation will be quite different: six vitally important genes of DNA metabolism and two ribosomal operons occur near oriC, and no vitally important genes of any category occur in the 6° region around terC. The difference is even more striking if we consider the very small 12-kb site (1°, or 0.3% of the chromosome length) in the beginning of the right replichore: this site harbors four vitally important genes of DNA metabolism (*dnaA, N; gyrB, A*) and one of the ribosomal operons. The absence of vitally important genes in chromosomal region around terC was shown in the 1980s by the studies in which mutants with deletions of this region were obtained [45, 46]. Thus, in a study with *E. coli* [45], a mutant had a 340-kb deletion that included terC. The colonies of this mutant grew slowly, and many of its cells produced filamentous with impaired division; however, it was viable and evidently did not lack genes deserving inclusion in the Kobayashi list. Moreover, a study of *B. subtilis* mutants [46] with large deletions (up to 230 kb) in the terC region did not reveal any decrease in the growth rate in rich medium.

The role of sites adjacent to terC in chromosome segregation was already mentioned above. It is possible that the increased recombination rate in this region did not allow vitally important genes to be evolutionarily fixed in this region: the constant rearrangement of DNA sites here would have led to too high lethality. Interestingly, the homology in this region between *E. coli* and other enterobacteria is lower than the homology in other regions, suggesting its high variability and rapid divergence [47]. More significant divergence of terC regions compared to oriC is noted in [48]. Another reason for the absence of vitally important genes in the terC region may be their delayed duplication in the course of chromosome replication and thus delayed duplication of the amount of their products.

It has been speculated that the terC region is a convenient zone for incorporation of foreign DNA during lateral gene transfer, since in this region, the recombination frequency is sharply increased. Indeed, one of the clusters of preferable incorporation of lambdoid prophages in the *E. coli* chromosome is not far from this region (see review [49]). In *B. subtilis*, a large region occupied by prophage SPβ genes occurs 30 kb to the left of terC. The terC region contains quite a large number of sequences enriched in AT pairs, whose presence has been attributed to the phage origin of these sequences [40]. However, the concentration of DNA of foreign origin is not so pronounced, and one can speak only about a tendency [1]. Thus, genes of some defec-

tive prophages in the *B. subtilis* chromosome are far from terC; for example, the genes of the defective prophage PBSX are in the middle of the right replichore (see review [50]).

It is appropriate to discuss here not only the tendencies in the gene disposition in the chromosome but also tendencies in the disposition of certain nucleotides fulfilling specific functions, namely, Chi and Rag sites. Chi sites feature rare octamers exhibiting a GCTGGTGG motif, whose conserved portion is TGGT (variations may occur). Chi sites promote homologous recombination, which is known to involve the Rec BCD enzyme systems. During recombination, structures resembling the Greek letter chi are formed, and this determined the name of these sequences. On average, each 5 kb of the *E. coli* genome contains one Chi site. The density of the occurrence of Chi sites changes along the replichore, increasing toward terC and peaking in the dif-active zone [51].

Rag sites are purine-rich octamers with a RRNAGGGS motif (R are purines, N means any base, and S is guanine or cytosine). The distribution of Rag sites over the replichores is even more polar than that of Chi sites; the concentration of Rag sites sharply increases in the dif-active zone [51, 52]. As mentioned above, recombinations and rearrangements of different kinds are especially frequent in this zone. The presence of the Rag octamers is apparently needed for recombination events and for repair of the resulting impairments of the chromosome.

## LOCATION OF GENES IN THE LEADING AND LAGGING DNA STRANDS

It has been mentioned above that the transcription of the coding sequences located on the lagging DNA strand may be impaired due to the fact that DNA polymerase III and RNA polymerase move in different directions along this strand. This is a prerequisite for preferable location of genes in the leading DNA strand. It would have been more precise to speak about the location of coding sequences of genes, but, for the sake of brevity, I use the commonly accepted terminology. It was before the genome sequencing epoch that the sequencing of individual genes revealed a tendency toward gene location in the leading DNA strand. This tendency was observed in *E. coli* and *B. subtilis* [43, 54]. Later, after complete sequencing of a great many genomes, this tendency was confirmed, although it is no more than a tendency. Thus, Rocha [55] analyzed 59 genomes of different bacterial species and found that, in *Firmicutes*, 72–82% of genes were located in the leading strand (in *B. subtilis*, 74%). However, in some representatives of this group and in other bacteria examined, this value was lower, 52–60% (54% in *E. coli*). Interestingly, the same tendency persisted in bacteria possessing more than one chromosome (*V. cholerae, B. melitensis*). Further analyses narrowed the range of the genes examined, limiting it to genes with a high expression level (judging from the so-called codon adaptation index) or to vitally important genes. It was found that genes of these two categories show strong preference to the leading strand. The two categories overlapped in part; e.g., both included genes of ribosomal operons. The choice of the leading strand is determined, most probably, by the gene vital importance 94% of the genes from the Kobayashi list proved to be located in the leading strand of the *B. subtilis* chromosome. A similar gene-location pattern was found in *E. coli*, although its genes on the whole show much weaker preference to the leading strain than *B. subtilis* genes [1, 57]. It is noteworthy that in *E. coli*, Chi and Rag sites are also located mainly in the leading strand [52].

An issue close to the scope of this review is the research line in computer genomics that explores such details of the bacterial genome structure as the purine/pyrimidine ratio in different sites of the leading and lagging DNA strands, synonymous substitutions of nucleotides, distribution of restriction sites, etc. [1, 51, 57]. Without considering these investigations in detail, I would like to mention one tendency that they revealed: and increase in guanine and a decrease in cytosine content with the distance from oriC and approach to terC [58]. This tendency is so pronounced that it served as the basis for a method for determination of the oriC location from the data on the complete nucleotide sequence of a genome. Thus, Oriloc software [59] precisely finds the oriC location in *B. burgdorferi* and *B. subtilis*, and the error in the case of *E. coli* is only 7 kb. In a purely in silico study [60], the existence of multiple replication origins in archaebacteria was predicted, which is known for eukaryotes. The basis in this tendency in guanine and cytosine distribution and its biological meaning (if any) remain unknown. One of the consequences of this tendency may be the uneven distribution of mutation frequency along the chromosome [61]. However, the study [62], in which the frequency of reversions in the *lacZ* gene inserted in different sites of *S. enterica* chromosome was determined, failed to reveal any correlation between changes in the reversion rate and the distance from oriC.

## IS GENE-LOCATION PATTERN IN THE CHROMOSOME AN ADAPTIVE TRAIT?

The bacterial cell can regulate, directly or indirectly, the amount of products of particular genes. This is achieved, e.g., via the operon organization of the genome and gene amplification. The polarity of the distribution of vitally important genes is probably one such mechanism. Most of these genes are located near the origin of replication, and, due to this, their number and, hence, the amount of their products, which are vitally important for the cell, are duplicated in the first turn. The order in which the amounts of gene products are duplicated may play a regulatory role in cases were

launching of a particular process requires a threshold amount of a certain protein (e.g., initiation of replication and the *dnaA* gene). The preferential location of vitally important genes in the leading DNA strand promotes the formation of their full-value transcripts, since genes with these location are safe from the consequences of the collisions of DNA and RNA polymerases. Vitally important genes are not subjected to the impairments inevitable in the segregation region, which result from decatenation and segregation of circle chromosomes. It can be said the vitally important genes are "the gold fund" of the bacterial cell and are under special protection.

All the more interesting are the exceptions to the above-considered rules. In *Mycobacterium tuberculosis*, an extremely slowly growing microorganism, the single ribosomal operon occurs very far from oriC [39]. What should be considered to be the reason of this chromosomal organization and what should be regarded as its consequence? Probably, here we deal with a phenomenon similar to reduction of an unused organ, a process well studied by comparative anatomy: reduction of most of ribosomal operons may have occurred because they were excessive under conditions of slow growth, determined by some other reasons. Alternatively, it may be assumed that slow growth of the tuberculosis causative agent is a property useful to it (although usually it is believed that rapid growth and multiplication should be favorable for a bacterium). It cannot be ruled out that, during a chronic disease such as tuberculosis, an extremely low multiplication rate of the causative agent is one of the prerequisites for long-term coexistence of the parasite (mycobacterium) and the host (human). For a parasite, it may be more profitable to behave more like a symbiont than like a predator. If this is so, then the low number of ribosomal genes in the chromosome of *M. tuberculosis* may be considered as the primary reason for its slow growth, and this peculiarity of the mycobacterial genome is an adaptive trait of selective importance.

## ACKNOWLEDGMENTS

I am grateful to E.U. Poluektova and D.V. Mukha for valuable consultations and to I.P. Shilovskii for technical assistance in the preparation of the manuscript.

This work was supported in part by a the program "Origin and Evolution of the Biosphere" of the Presidium of the Russian Academy of Sciences.

## REFERENCES

1. Rocha, E., The Replication-Related Organization of Bacterial Genomes, *Microbiology* (UK), 2004, vol. 150, no. 5, pp. 1609–1627.
2. Prozorov, A.A., Bacterial Genome Organization: Uniformity or Diversity? *Genetika,* 1995, vol. 31, no. 6, pp. 741–752.
3. Fonstein, M. and Haselkorn, R., Physical Mapping of Bacterial Genome, *J. Bacteriol.*, 1995, vol. 117, no. 17, pp. 3361–3369.
4. Prozorov, A.A., The Bacterial Genome: Nucleoid, Chromosome, and Nucleotide Map, *Mikrobiologiya*, 1998, vol. 67, no. 4, pp. 437–451 [*Microbiology* (Engl. Transl.), vol. 67, no. 4, pp. 361–372].
5. Casjens, Sh., The Diverse and Dynamic Structure of Bacterial Genomes, *Annu. Rev. Genet.*, 1998, vol. 32, pp. 339–377.
6. Yoshikawa, H. and Ogasawara, N., Structure and Function of DNA A and dnaA Boxes in Eubacteria: Evolution of Bacterial Replications Origins, *Mol. Microbiol.*, 1991, vol. 5, no. 11, pp. 2589–2597.
7. Zawilac-Pawlik, A., Kois, A., Majaka, J., Jakomowic, D., Smulczuk-Krawczyszun, A., Messer, W., and Zakrzewska-Czerwinska, J., Architecture of Bacterial Replications Initiations Complexes: Orisomes from Four Unrelated Bacteria, *Biochem. J.*, 2005, vol. 389, no. 3, pp. 471–481.
8. Messer, W., The Bacterial Replications Initiator DnaA. DnaA and oriC, the Bacterial Mode To Initiate DNA Replication, *FEMS Microbiol. Rev.*, 2002, vol. 26, pp. 355–374.
9. Yuzhakov, A., Turner, J., and O'Donnell, M., Replisome Assembly Reveals the Basis for Asymmetric Function in Leading and Lagging Strand Replication, *Cell*, 1996, vol. 86, no. 6, pp. 877–886.
10. Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapius, J., Errington, J., Janniere, L., and Ehrlich, D., Two Essential DNA Polymerases at the Bacterial Replication Fork, *Science*, 2001, vol. 294, no. 5546, pp. 1716–1718.
11. Liu, B. and Alberts, B., Head-on Collision between a DNA Replication Apparatus and RNA Polymerase Transcription Complex, *Science*, 1995, vol. 267, no. 5201, pp. 1131–1137.
12. Francino, M.P. and Ochma, H., Strand Asymmetries in DNA Evolution, *Trends Genet.*, 1997, vol. 13, no. 6, pp. 240–245.
13. Ellwood, M. and Nomura, M., Chromosomal Locations of the Genes for rRNA in *Escherichia coli* K-12, *J. Bacteriol.*, 1982, vol. 149, no. 2, pp. 458–468.
14. Russel, D., Hemimethylation Prevent DNA Replication in *E. coli, Cell*, 1987, vol. 50, no. 7, pp. 1071–1079.
15. Liu, M., Campbell, J., and Kleckner, N., SeqA: a Negative Modulator of Replication Initiation in *E. coli, Cell*, 1994, vol. 77, no. 3, pp. 413–426.
16. Yoshikawa, H. and Sueoka, N., Sequential Replication of *Bacillus subtilis* Chromosome. 1. Comparison of Marker Frequencies in Exponential and Stationary Growth Phases, *Proc. Natl. Acad. Sci. USA*, 1963, vol. 9, no. 4, pp. 559–566.
17. Prozorov, A.A., Recombinational Rearrangements in Bacterial Genome and Bacterial Adaptation to the Environment, *Mikrobiologiya*, 2001, vol. 70, no. 5, pp. 581–594 [*Microbiology* (Engl. Transl.), vol. 70, no. 5, pp. 501–511].
18. Chandler, M. and Pritchard, R., The Effect of Gene Concentration and Relative Gene Dosage on Gene Output in *Escherichia coli, Mole. Gen. Genet.*, 1975, vol. 138, no. 2, pp. 127–141.

19. Schmid, M.. and Roth, J.R., Gene Location Affects Expression Level in *Salmonella typhimurium, J. Bacteriol.*, 1987, vol. 169, no. 6, pp. 2872–2875.
20. Prozorov, A.A., The Bacterial Cell Cycle: DNA Replication, Nucleoid Segregation, and Cell Division, *Mikrobiologiya*, 2005, vol. 74, no. 4, pp. 437–451 [*Microbiology* (Engl. Transl.), vol. 74, no. 4, pp. 375–387].
21. Kamada, K., Horiushi, T., Ohsumi, K., Shimamoto, N., and Moricawa, K., Structure of a Replication-Terminator Protein Complexed with DNA, *Nature*, 1996, vol. 383, no. 6601, pp. 598–603.
22. Blakely, G., May, G., Mc Culloch, R., Arciszewska, L., Burke, M., Lowett, S., and Sherratt, D., Two Related Recombinases Are Required for Site-Specific Recombination at dif and cer in *E. coli* K-12, *Cell*, 1993, vol. 75, no. 2, pp. 351–361.
23. Horiushi, T., Fujimura, J., Nishitani, H., Kobayashi, T., and Hidaka, M., The DNA Replication Fork Blocked at the ter Site May Be an Entrance for the Rec BCD Enzyme into Duplex DNA, *J. Bacteriol.*, 1994, vol. 176, no. 15, pp. 4656–4663.
24. Corre, J. and Louarn, J.-M., Evidence from Terminal Recombination Gradients That FtsK Uses Replichore Polarity to Control Chromosome Terminus Positioning at Division in *Escherichia coli, J. Bacteriol.*, 2002, vol. 184, no. 14, pp. 3801–3807.
25. Louarn, J.-M., Louarn, J., Fransois, V.., and Patte, J., Analysis and Possible Role of Hyperrecombination Region of the *Escherichia coli* Chromosomes, *J. Bacteriol.*, 1991, vol. 173, no. 16, pp. 5097–5104.
26. Hill, T.M, Feature of the Chromosomal Terminus Region, *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed., Neighardt, F.C., Ed., Washington, DC: ASM, 1996, pp. 1602–1614.
27. Condon, C., French, S., and Squires, C.L., Deletion of Functional Ribosomal RNA Operons in *Escherichia coli* Causes Increased Expression of the Remaining Intact Copies, *EMBO J.*, 1993, vol. 12, no. 11, pp. 4305–4315.
28. Bremer, H and Dennis, P.P., Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed., Neighardt, F.C., Ed., Washington, DC: ASM, 1996, pp. 1553–1569.
29. De Martelaere, D. and Van Gool, A., The Density Distribution of Gene Loci Over the Genetic Map of *Escherichia coli*: Its Structural, Functional and Evolutionary Implication, *J. Mol. Evol.*, 1981, vol. 17, no. 6, pp. 354–360.
30. Sukhodolets, V.V., Organization of the *Escherichia coli* Genome, *Genetika*, 1985, vol. 21, no. 5, pp. 693–706.
31. Sukhodolets, V.V., Principles of the Prokaryotic Genome Organization, *Genetika*, 1992, vol. 28, no. 1, pp. 28–37.
32. Zakharov, I.A. and Matselyukh, B.P., *Geneticheskie karty mikroorganizmov* (Genetic Maps of Microorganisms), Kiev: Naukova dumka, 1986, pp. 56–58.
33. Prozorov, A.A., The Genome of Bacterial Cell, *Organizatsiya genoma* (Genome Organization), Moscow: Nauka, 1989.
34. Burnland, V., Plunkett, G., Daniels, D., and Blattner, F., DNA Sequence and Analysis of 136 Kilobases of the *Escherichia coli* Genome: Organizational Symmetry around the Origin of the Replication, *Genomics*, 1993, vol. 16, no. 3, pp. 551–561.
35. Ogasawara, N., Seiki, M., and Yoshikawa, H., Replication Origin Region of *Bacillus subtilis* Chromosome Contains Two rRNA Operons, *J. Bacteriol.*, 1983, vol. 154, no. 1, pp. 50–57.
36. Lewis, P., Thaker, Sh.., and Errington, J., Compartmentalization of Transcription and Translation in *Bacillus subtilis, EMBO J.*, 2000, vol. 19, no. 4, pp. 710–718.
37. Dawies, K.M., and Lewis, P., Localization of rRNA Synthesis in *Bacillus subtilis*: Characterization of Loci Involved in Transcription Focus Formation, *J. Bacteriol.*, 2003, vol. 185, no. 7, pp. 2346–2353.
38. Aiyar, G., Gaal, R., and Gourse, R., rRNA Promoter Activity in the Fast-Growing Bacterium *Vibrio natriegens, J. Bacteriol.*, 2002, vol. 184, no. 5, pp. 1349–1358.
39. Cole, S.T., Brosch, R., Parkhill, J., Garniel, T., et al. (62 authors), Deciphering the Biology of *Mycobacterium tuberculosis* from the Complete Genome Sequence, *Nature*, 1998, vol. 393, no. 6685, pp. 537–544.
40. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., et al. (151 authors), The Complete Genome Sequence of the Gram-Positive Bacterium *Bacillus subtilis, Nature*, 1997, vol. 390, no. 6675, pp. 249–256.
41. Kobayashi, K., Ehlich, D., Albertini, A., Amati, G., et al. (99 authors), Essential *Bacillus subtilis* Genes, *Proc. Natl. Acad. Sci. USA*, 2003, vol. 100, no. 8, pp. 4678–4683.
42. Vagner, V.., Dervyn, E., and Ehrlich, S.D., A Vector for Systematic Gene Inactivation in *Bacillus subtilis, Microbiology* (UK), 1998, vol. 144, no. 8, pp. 3097–3104.
43. Ogasawara, N. and Yoshikawa, H., Genes and Their Organization in the Replication Origin of the Bacterial Chromosome, *Mol. Microbiol.*, 1992, vol. 6, no. 3, pp. 629–634.
44. Old, I., Margarita, D., and Saint GironsI., Unique Genetic Arrangement in the dnaA Region of the *Borrelia burgdorferi* Linear Chromosome: Nucleotide Sequence of the *dnaA* Gene, *FEMS Microbiol. Lett.*, 1993, vol. 111, no. 1, pp. 109–114.
45. Henson, J., and Kuempel, P., Deletion of the Terminus Region (340 kilobase pairs of DNA) from the Chromosome of *Escherichia coli, Proc. Natl. Acad. Sci. USA*, 1985, vol. 82, no. 11, pp. 3766–3770.
46. Lismaa, T. and Wake, R., The Normal Replication Terminus of the *Bacillus subtilis* Chromosome, terC, Is Dispensable for Vegetative Growth and Sporulation, *J. Mol. Biol.,* 1987, vol. 195, no. 2, pp. 299–310.
47. Carp, P., Shields, D., Wolfe, K., and Wen-Hsiung Li, Chromosomal Location and Evolutionary Rate Variation in Enterobacterial Genes, Science, 1989, vol. 246, no. 4930, pp. 808–810.
48. Mira, A. and Ochman, H., Gene Location and Bacterial Sequence Divergence, *Mol. Biol.*, 2002, vol. 19, no. 6, pp. 1350–1358.
49. Campbell, A., Chromosomal Insertion Sites for Phages and Plasmids, *J. Bacteriol.*, 1992, vol. 174, no. 23, pp. 7495–7499.
50. Prozorov, A.A., Defective Phages of Bacilli: Cell Parasites or Chromosomal Components? *Genetika*, 1996, vol. 32, no. 4, pp. 469–481 [*Russ. J. Gen. Genet.* (Engl. Transl.), vol. 32, no. 4, pp. 403–414].

51. Lobry, J. and Louarn, J.-M., Polarization of Prokaryotic Chromosomes, *Curr. Opin. Microbiol.*, 2003, vol. 6, pp. 101–108.
52. Capiaux, H., Cornet, F., Guijo, M-I., Corre, J., Guijo, M.-I., Perals, K., Rebollo, J., and Louarn, J.M., Polarization of the *Escherichia coli* Chromosome. A View from the Terminus, *Biochemi,* 2001, vol. 83, no. 1, pp. 161–170.
53. Brewer, B., When Polymerases Collide: Replication and the Transcriptional Organization of the *E. coli* Chromosome, *Cell*, 1988, vol. 83, no. 3, pp. 679–686.
54. Ziegler, D. and Dean, D., Orientation of Genes in the *Bacillus subtilis* Chromosome, *Genetics*, 1990, vol. 125, no. 4, pp. 703–708.
55. Rocha, E., Is There a Role for Replication Fork Asymmetry in the Distribution of Genes in Bacterial Genomes? *Trends in Microbiology*, 2002, vol. 10, no. 9, pp. 393–395.
56. Sharp, P.M. and Li, W., The Codon Adaptation Index—a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications, *Nucleic Acids Res.*, 1987, vol. 15, no. 6, pp. 1281–1295.
57. Rocha, E., Viari, A., and Danchin, A., Oligonucleotide Bias in *Bacillus subtilis*: General Trends and Taxonomic Comparisons, *Nucleic Acids Res.*, 1998, vol. 26, no. 12, pp. 2972–2980.
58. Tiller, E. and Collins, R., Replication Orientation Affects the Rate and Direction of Bacterial Gene Evolution, *J. Mol. Evol.*, 2000, vol. 51, no. 2, pp. 459–463.
59. Frank, A. and Lobry, J., Oriloc: Prediction of Replication Boundaries in Unannotated Bacterial Chromosomes, *Bioinformatics*, 2000, vol. 16, no. 6, pp. 560–561.
60. Nikoelaou, Ch. and Almirantis, Y., A Study on the Correlation of Nucleotide Skews and the Positioning of the Origin of Replication: Different Modes of Replication in Bacterial Species, *Nucleic Acids Res.*, 2005, vol. 33, no. 21, pp. 6816–6822.
61. Sharp P., Shields K., Wolfe K, and Li, W., Chromosomal Location and Evolutionary Rate Variation in Enterobacterial Genes, *Science*, 1982, vol. 246, no. 4, pp. 808–810.
62. Hudson, R., Bergthorsson, U., Roth, J., and Ochman, H., Effect of Chromosome Location on Bacterial Mutation Rates, *Mol. Biol. Evol.*, 2002, vol. 19, no. 1, pp. 85–92.