ISSN 1022-7954, Russian Journal of Genetics, 2006, Vol. 42, No. 11, pp. 1303–1312. © Pleiades Publishing, Inc., 2006.
Original Russian Text © A.V. Markov, I. A. Zakharov, 2006, published in Genetika, 2006, Vol. 42, No. 11, pp. 1547–1557.

Large and Small Rearrangements in the Evolution of Prokaryotic Genomes

A. V. Markov[a] **and I. A. Zakharov**[b]

[a] *Paleontological Institute, Russian Academy of Sciences, Moscow, 117997 Russia; e-mail: markov_a@inbox.ru*
[b] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia; e-mail: zakharov@vigg.ru*

Received February 20, 2006

Abstract—Relative frequencies of large and small genome rearrangements (inversions and transpositions) in the evolution of prokaryotic genomes can be evaluated using the ratio between the index S (the ratio of the number of identical pairs of neighboring genes in two genomes to the total number of genes in the sample of interest) and $1 - 6L/n$, where L is the mean difference in intergenic distances and n is the number of genes in the sample. The S value uniformly decreases with the fixation of genome rearrangements, while the decrease rate of $1 - 6L/n$ is determined by the rearrangement size. Specifically, large inversions and transpositions lead to a dramatic decrease in the index value, while small rearrangements result in an insignificant decrease. The ratio between these indices was computed for twenty pairs of closely related species belonging to different groups of bacteria and archaea. The pairs examined strongly differed in the relative frequency of large and small rearrangements. However, computer simulation showed that the total variation can be reproduced with the same input parameters of the model. This means that the differences observed can be stochastic and can be interpreted without assuming different mechanisms and factors of genome rearrangements for different groups of prokaryotes. Relative frequencies of large and small rearrangements displayed no noticeable correlations with taxonomic position, total rate of rearrangement fixation, habitation conditions, and the abundance of transposons and repetitive sequences. It is suggested that, in some cases, phage activity increases the frequency of large genome rearrangements.

DOI: 10.1134/S1022795406110123

INTRODUCTION

In recent years, patterns of evolutionary changes in gene order in chromosomes of different organisms have attracted increasing attention of researches. The gene order changes as a result of inversions, translocations, and transpositions. One of us [1] was the first to suggest quantitative measuring of the similarity in gene order on genetic maps. Based on the similarity measures suggested, a comparison of mammalian genomes was done, and the possibility of reconstructing the genome phylogeny was demonstrated [2, 3]. Furthermore, evolution of genomes consisting of several chromosomes was simulated [4, 5]. The similarity measures were designed for comparing genomes with an unknown gene order in the chromosomes and a known assignment of the genes to certain chromosomes. The measure of similarity was also suggested for gene orders themselves, which allowed a comparison of genomes represented not by synteny groups (in which the gene orders were unknown) but by linkage groups [6].

Another measure for comparisons of this type was independently developed by Sankoff and colleagues, who compared the gene orders in mitochondria from different organisms [7, 8]. In further investigations, these authors examined the gene order changes in mammalian chromosomes [9].

As the number of prokaryotic genome sequences increased, different quantitative and qualitative methods of gene order comparisons were suggested and applied for the analysis of prokaryotic genomes [10–17]. One of the main conclusions of these and other studies was that the gene order similarity was generally in good correlation with the phylogenetic relationships of the species (strains) compared. These findings provided for the use of quantitative estimates of this similarity to construct meaningful dendrograms by use of standard methods (Fitch–Margoliash, neighbor-joining, and others). The resulting trees, along with those based on nucleotide sequence similarity between individual genes, can be used in phylogenetic studies. The programs for phylogenetic reconstructions based on gene order comparisons have been developed and are freely available from the Internet (SHOT vs. 2.0, Shared Ortholog and Gene Order Tree Reconstruction Tool: http://www.bork.embl-heidelberg.de/SHOT/) [18].

At the same time, gene order evolution in different groups of prokaryotes has its specifics, which should be considered while interpreting the results of comparative genome analyses. Two of these specific features, in which the patterns of gene order evolution can substantially differ in different prokaryotes, should be mentioned. These are the rate (frequency) of rearrangements and their size (the ratio between small and large rearrangements).

Remarkable differences in the fixation rates of genome rearrangements in different groups were reported in a number of publications. For instance, an abnormally high rate of gene order changes is typical of *Wolbachia* bacteria, intracellular parasites of terrestrial invertebrates [19]. An increased number of genome rearrangements was also observed by comparing *Pyrococcus* and *Mycobacterium* species. At the same time, a limited number of rearrangements was detected in *Mycoplasma*, *Chlamydia*, *Helicobacter*, and some other prokaryotes [14, 20–28].

The problem of the ratio between large and small genome rearrangements in the evolution of different prokaryotic groups is poorly understood. The available literature data are controversial. For instance, a predominance of large inversions was described for prokaryotes as opposed to *Saccharomyces cerevisiae* and *Candida albicans* fungi, where half of the inversions were small and included from one to several genes [29]. At the same time, there are reports that small inversions prevail in prokaryotes, contrary to animals and plants, in which the prevalence of small inversions is not so high [30, 31]. Sankoff and colleagues attracted attention to the fact that the quantitative ratio between large and small genome rearrangements can lead to quantitative differences in the levels of conservation of long conserved segments with identical gene orders and gene clusters, which are groups of neighboring genes with different orders within the group. In the case of a predominance of small rearrangements, the relatedness of the organisms compared can be inferred from the number of shared gene clusters. In the opposite case, the number of conserved segments is more demonstrative [31]. Several complicated formulas were suggested to calculate the probability of a random occurrence of a certain number of gene clusters with different characteristics within two completely shuffled genomes as a function of the number of genes in the cluster (m); genome size (n); and the window size (r), i.e., the length of the genome region within which the desired m genes should be found [32, 33].

Gene order comparisons performed in α-proteobacteria suggested an increased frequency of small genome rearrangements relative to large rearrangements [19], although the probability of large rearrangements in this bacterial group was rather high.

In the current study, we developed a new method to quantitate the ratio between large and small genome rearrangements in the evolution of different groups of prokaryotes, based on a comparison of the number of breakpoints and the mean difference in intergenic distances. Based on the gene order comparisons in pairs of closely related species from different prokaryotic groups, we estimated the differences between these species by the relative frequencies of large and small genome rearrangements and tried to reveal the possible causes of these differences.

MATERIALS AND METHODS

Genomes of 34 prokaryotic species were analyzed (Table 1). Gene order comparisons were performed in twenty pairs of closely related species (Table 2). These genome pairs were chosen based on preliminary gene order comparisons (pairs with medium and high numbers of identical neighboring gene pairs were taken into analysis). The set of homologous genes examined was the same as in a previous study [19]. The exception was the ribosomal protein genes, forming the so-called ribosomal superoperon. A specific feature of these genes is that, during evolution, they have preserved a similar order even in unrelated groups of prokaryotes (their order is partly conserved even in the plastid and mitochondrial genomes) [34–36]. However, it should be noted that exclusion of the ribosomal protein genes from the sample analyzed had only a weak effect on the results.

A table of orthologous genes and their order in chromosomes are available at http://macroevolution.narod.ru/bact.htm (bacteria) and http://macroevolution.narod.ru/arch.htm (archaea).

Quantitative measures used were as follows [19].

(1) The relative number of neighboring gene pairs shared by two genomes was estimated as

$$S = m/n,$$

where m is the number of identical pairs of neighboring genes in two genomes compared (the genes were considered as neighbors when there were no genes from the sample examined between them on the chromosome) and n is the number of homologous genes used in comparing the genomes of interest. The S value reflects the degree of similarity in the exact gene order in two genomes compared (i.e., the degree of conservation of conserved segments).

(2) The mean difference in the distances between genes, L, was calculated as follows. For each gene, the distances to the other genes in the genome examined were calculated. In total, $(n^2 - n)/2$ intergenic distances were calculated for each gene. Each intergenic distance was calculated as an absolute difference between the conventional position numbers of two genes in the genome (the conventional number of a position was defined as the ordinal number of a gene in the sequence of n genes examined). Considering that the chromosome was circular, an intergenic distance x higher than $n/2$ was replaced by $n - x$. Then, the absolute difference in intergenic distances was calculated for each of the $(n^2 - n)/2$ pairs of homologous genes in two genomes to be compared. The resultant L value was calculated as the mean of these absolute values.

To compare the data obtained for different genome pairs with different n values, normalized value $6L/n$ was used (model experiments showed that, with accumulation of genome rearrangements, L converged to $n/6$ at any n). This value, in turn, was subtracted from unity to obtain the similarity measure (instead of differ-

Table 1. Prokaryotic genomes used in the study

Species	Strain	Genome size, kb	Taxonomy	Desig-nation	GenBank accession no.
Wolbachia pipientis	Endosymbiont of *D. melanogaster*	1267782	Alphaproteobacteria, Rickettsiales, Wolbachieae	wMe	NC_002978
W. sp.	Endosymbiont of *Brugia malayi* str. TRS	1080084	–	wBm	NC_006833
Rickettsia conorii	str. Malish 7	1268755	Alphaproteobacteria, Rickettsiales, Rickettsieae	ric	NC_003103
R. prowazekii	str. Madrid E	1111523	–	rip	NC_000963
R. felis	URRWXCal2	1485148	–	rif	NC_007109
Anaplasma marginale	str. St. Maries	1197687	Alphaproteobacteria, Rickettsiales, Anaplasmataceae	ana	NC_004842
Ehrlichia ruminantum	str. Welgevonden	1512977	–	her	NC_006832
Escherichia coli	CFT073	5231428	Gammaproteobacteria, Enterobacteriales, Enterobacteriaceae	eco	NC_004431
Salmonella enterica subsp. *enterica*	Serovar. Typhi Ty2	4791961	–	sal	NC_004631
Shigella flexneri	2a str. 301	4607203	–	shg	NC_004337
Yersinia pestis	KIM	4600755	–	yep	NC_004088
Pasteurella multocida subsp. *multocida*	Pm70	2257487	Gammaproteobacteria, Pasteurellales, Pasteurellaceae	pas	NC_002663
Chlamydia trachomatis	D/UW-3/CX	1042519	Chlamydiae, Chlamydiales, Chlamydiaceae	clt	NC_000117
Chlamydophila pneumoniae	AR39	1229853	–	cpp	NC_002179
Corynebacterium glutamicum	ATCC 13032	3309401	Actinobacteria, Actinobacteridae, Actinomycetales, Corynebacterineae, Corynebacteriaceae	cor	NC_003450
Mycobacterium leprae	TN	3268203	Actinobacteria, Actinobacteridae, Actinomycetales, Corynebacterineae, Mycobacteriaceae	myl	NC_002677
M. tuberculosis	CDC1551	4403837	–	myt	NC_002755
Synechococcus sp.	WH 8102	2434428	Cyanobacteria, Chroococcales	syn	NC_005070
Prochlorococcus marinus	MIT 9313	2410873	Cyanobacteria, Prochlorales, Prochlorococcaceae	pro	NC_005071
Clostridium perfringens	13	3031430	Firmicutes, Clostridia, Clostridiales, Clostridiaceae	clp	NC_003366
C. acetobutylicum	ATCC 824	3940880	–	cla	NC_003030
Streptococcus pneumoniae	R6	2038615	Firmicutes, Lactobacillales, Streptococcaceae	spn	NC_003098
S. pyogenes	SSI-1	1894275	–	spy	NC_004606
Bacillus halodurans	C-125	4202352	Firmicutes, Bacillales, Bacillaceae	bah	NC_002570
B. subtilis subsp. *subtilis*	168	4214630	–	bas	NC_000964
Listeria innocua	Clip11262	3011208	Firmicutes, Bacillales, Listeriaceae	lii	NC_003212
Methanosarcina acetivorans	C2A	5751492	Euryarchaeota, Methanomicrobia, Methanosarcinales, Methanosarcinaceae	msa	NC_003552
M. mazei	Go1	4096345	–	msm	NC_003901
Thermoplasma acidophilum	DSM 1728	1564906	Euryarchaeota, Thermoplasmata, Thermoplasmatales, Thermoplasmataceae	tpa	NC_002578NC_002578
T. volcanicum	GSS1	1584804	–	tpv	NC_002689
Pyrococcus furiosus	DSM 3638	1908256	Euryarchaeota, Thermococci, Thermococcales, Thermococcaceae	pyf	NC_003413
P. horikoshii	OT3	1738505	–	pyh	NC_000961
Sulfolobus solfataricus	P2	2992245	Crenarchaeota, Thermoprotei, Sulfolobales, Sulfolobaceae	sus	NC_002754
S. tokodaii	7	2694756	–	sut	NC_003106

Table 2. Comparison of gene orders in the pairs of closely related prokaryotic species

Pair	*n*	*S*	*L*	1–6*L*/*n*
wMe/wBm	155	0.426	25.4	0.017
ana/ehr	155	0.742	15.1	0.415
ric/rip	155	0.968	0.76	0.971
ric/rif	155	0.910	12.7	0.507
eco/sal	155	0.813	15.4	0.404
eco/shg	155	0.877	2.70	0.895
eco/yep	154	0.773	16.2	0.369
eco/pas	140	0.279	21.2	0.091
clt/cpp	102	0.706	9.33	0.451
myt/myl	113	0.664	15.4	0.182
myt/cor	110	0.727	13.9	0.242
syn/pro	123	0.805	14.5	0.293
cla/clp	110	0.573	12.9	0.296
spn/spy	104	0.327	13.4	0.227
bah/bas	124	0.782	12.9	0.376
bah/lii	122	0.598	17.7	0.130
msa/msm	160	0.875	4.74	0.822
tpa/tpv	159	0.667	15.6	0.411
pyf/pyh	159	0.616	13.6	0.487
sus/sut	160	0.806	22.6	0.153

Note: *n*, the number of homologous genes used for the comparison.

ence measure), comparable to the *S* index. The 1 – 6*L*/*n* value first of all reflects the degree of similarity in approximate gene order between the two genomes compared (i.e., the degree of conservation of gene clusters).

The ratio between the *S* and 1 – 6*L*/*n* values can be used as an estimate of relative frequency of large and small genome rearrangements that have appeared in the course of divergence of this genome pair. Model experiments showed that, with accumulation of genome rearrangements, *S* decreases more or less uniformly. It reflects the number rather than the size of rearrangements. As genome rearrangements accumulate, 1 – 6*L*/*n* also decreases, albeit not uniformly. The decrease rate strongly depends on the size of rearrangements. Large rearrangements result in dramatic decrease, while small rearrangements lead to only small decrease of this index.

CORRELATION BETWEEN *S* AND 1 – 6*L*/*n* VALUES IN EXPERIMENTS WITH SIMULATION OF GENOME REARRANGEMENTS

To reveal the relationships between the dynamics of the *S* and 1 – 6*L*/*n* indices and the relative frequency of large and small genome rearrangements, we used a computer model of genome order evolution, which was developed earlier [19]. The evolution of a circular genome containing 120 genes was simulated. In each experiment, 100 events (inversions and transpositions) were imitated. The probability for the next event to be inversion (but not transposition) was taken as 0.5 (i.e., the numbers of inversions and transpositions in the evolution of the model genome were nearly equal). This simulation seems to be in good correlation with the events taking place in the evolution of real prokaryotic genomes. For instance, during divergence of *Chlamydia trachomatis* and *C. pneumoniae*, there were 59 ± 6 events, 51% of which were inversions (predominantly short) and 49% were transpositions [30].

Model experiments showed the following.

(1) The *S* dynamics weakly depends on the rearrangement size; a decrease in *S* is, on the whole, proportional to the number of rearrangements. Transpositions lead to a more rapid (1.5 times) decline as compared to inversions, since each transposition creates three breakpoints, while inversion creates only two.

(2) The 1 – 6*L*/*n* dynamics strongly depends on the size of rearrangements. In the case of only small rearrangements, the decrease in this index is slow (Fig. 1). In all plots, the point (1, 1) corresponds to the complete identity of gene orders, while the point (0, 0) corresponded to maximum dissimilarity of the gene orders. As genomic differences accumulate, points shift leftward (reflecting a decline in *S*) and downward (decline in 1 – 6*L*/*n*). The 1 – 6*L*/*n* decline accelerates with increasing relative probability of large rearrangements (Figs. 2–4).

(3) Since the *S* dynamics depends only on the number of rearrangements, while the dynamics of 1 – 6*L*/*n* additionally depends on their sizes, the ratio between these indices can be used as a relative rate of large and small genome rearrangements occurring during the divergence in pairs of prokaryotic species. Introducing different probabilities of large and small rearrangements in the model, we obtained clearly shaped distribution areas of points on the plot (Figs. 1–4). These areas partly overlap, especially at maximum (close to 1) and minimum (close to 0) values of *S*. Nevertheless, within the wide range of intermediate *S* values, the position of a point on the plot characterizes the relative probability of large and small genome rearrangements.

(4) A rather wide scattering of points with the same parameters of the model (especially in the case where large rearrangements are allowed but are relatively rare; Fig. 3) indicates that, analyzing the course of evolution of model or real pairs of genomes, one should distinguish between relative probability of large rearrangements and their relative rate. The probability is determined only by the preset parameters of the model, while the real rate is determined by the model parameters in combination with the factor of randomness. For instance, when the probability of large rearrangements is equal to 0.1, this does not mean that every tenth rearrangement is a large one. In the model, the probabilities are preset artificially, via changing the model parame-

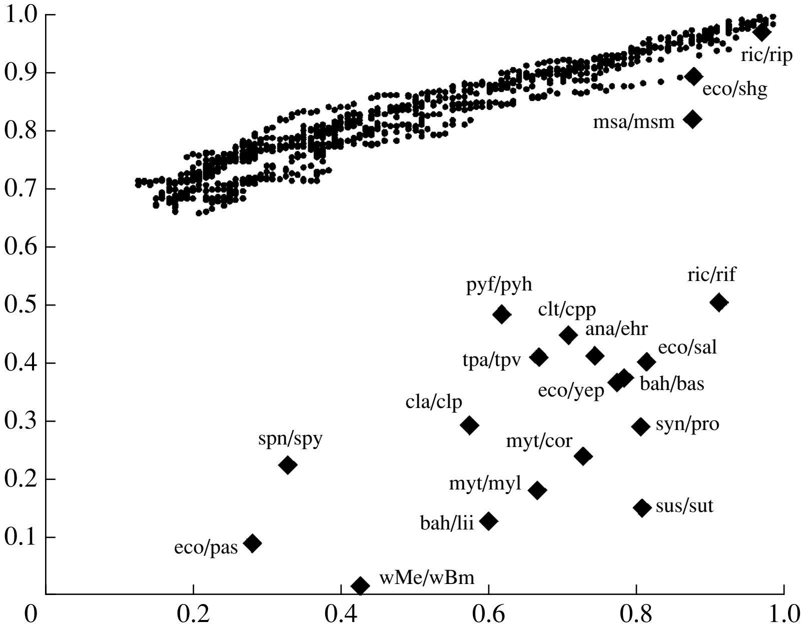

Fig. 1. The ratio between the *S* (abscissa) and 1 – 6*L*/*n* (ordinate) indices in real and model prokaryotic genome pairs. Model: only very small genome rearrangements are allowed (the excised segments are not longer than 6 genes; at transposition, they are transferred not further than by 6 positions). Ten executions of the model are presented. The model parameters: the size of the excised segment is 1 gene (the probability is 0.5), or 2–6 genes (0.5). Transposition distance: 0 (the probability is 0.5; in this case, inversion always occurs), 1 (0.2), 2–6 (0.3). Hereafter, the probability of insertion in the reverse order at transposition is 0.5; large symbols, real genome pairs; small symbols, model.

ters. In evolution of natural microorganisms, these parameters are determined by real molecular mechanisms of genome rearrangements (activity of phages and transposons, specific features of the replication enzymes, etc.), probably, in combination with environmental factors. However, in both cases (either model or natural conditions), these probabilities, having quite material bases, available for the analysis, determine not the exact frequencies of large and small rearrangements, but some possible (rather wide) range of frequencies. The width of this range can be evaluated from the point scatter in Figs. 1–4.

INTERPRETATION OF EMPIRICAL DATA ON 20 PAIRS OF PROKARYOTIC SPECIES BASED ON THE SIMULATION DATA

The pairs of closely related prokaryotic species examined remarkably differed from each other relative to the ratio between *S* and 1 – 6*L*/*n* (Fig. 1). In other words, in the course of evolution of these genome pairs, relative frequencies of large and small rearrangements substantially differed. In this context, there is a question of whether the data obtained are sufficient to make a conclusion that the scatter observed is caused by the differences in the mechanisms of genome rearrangements in different prokaryotes, or, in other words, that the differences in relative rates of large and small rearrangements reflect the differences in the corresponding probabilities.

The data from Fig. 3 clearly show that such a conclusion is baseless. With the same preset probabilities of large and small rearrangements, evolution of the twenty model genome pairs reproduced the whole range of the variation of the ratio between the *S* and 1 – 6*L*/*n* indices observed in twenty pairs of real prokaryotic species.

Thus, the data obtained did not conflict with the hypothesis that the ratio between probabilities of large and small rearrangements was identical during divergence of the twenty species pairs examined (moreover, large rearrangements were possible but less probable compared to small rearrangements). From here it follows that there are no grounds for the conclusion that the mechanisms and factors determining the sizes of rearrangements substantially differ in these species. It seems likely that the whole scatter observed can be explained by random events.

At the same time, the alternative hypothesis of substantially different probabilities of large and small rearrangements (and, consequently, different mechanisms and factors of genome rearrangements) is still not rejected. This assertion is supported by the results of simulation (Figs. 2, 4). Figure 2 shows that only some species pairs with the highest (relative to *S*) 1 – 6*L*/*n* values (pyf/pyh, spn/spy, eco/shg, msa/msm, and, to a lesser extent, clt/cpp and tpa/tpv) fit the hypothesis of an extremely low or null probability of large rearrangements (hypothesis 1). Figure 4 shows that a number of species pairs with minimum (relative to *S*) 1 – 6*L*/*n* val-

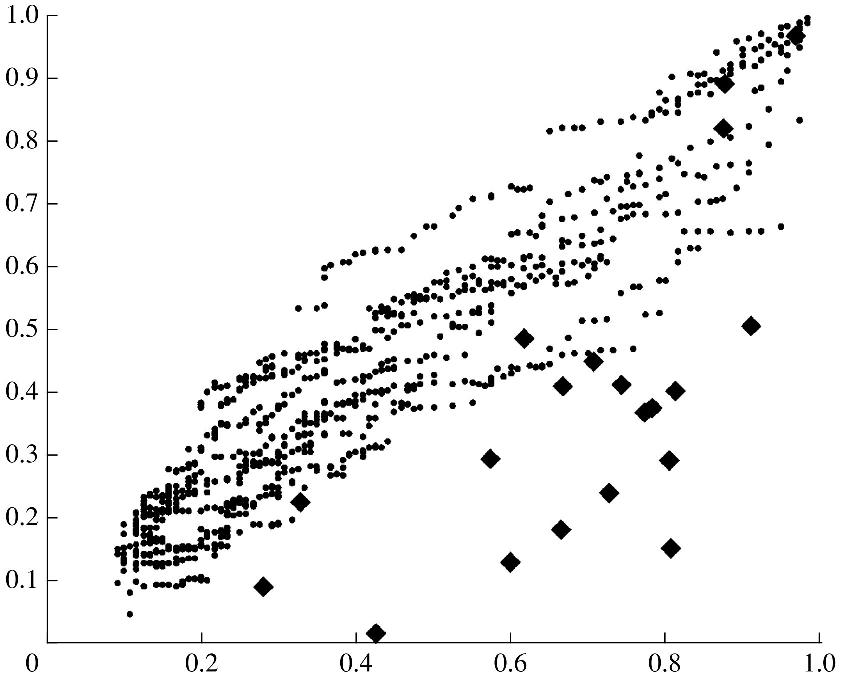

Fig. 2. Model: only small and medium rearrangements are allowed (the excised segments are not longer than 6 genes; at transposition, they are transferred to any distance). Ten executions of the model are presented. The model parameters: the size of the excised segment is 1 gene (the probability is 0.15), or 2–6 genes (0.5). Transposition distance: 0 (the probability is 0.5), 1 (0.02), 2–6 (0.08), 7–14 (0.2), or 15–60 (0.2). Here and in Figs. 3 and 4, the genome pairs are designated as in Fig. 1.

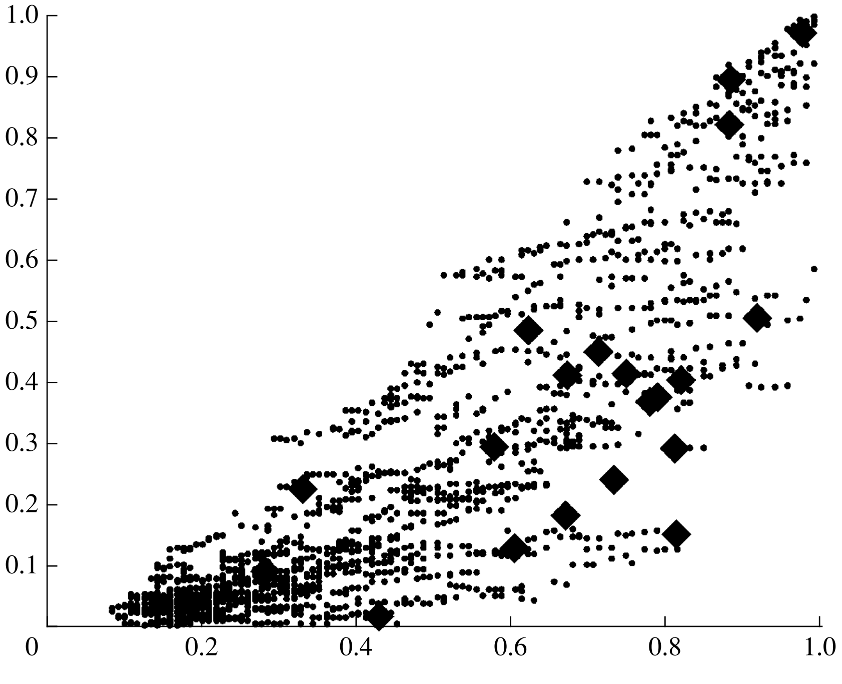

Fig. 3. Model: the probability of small rearrangements is higher than that of large ones (the segments longer than 6 genes are excised in 15% of the cases, and those longer than 14 genes are excised in 5% of the cases; at transposition, the segments are transferred at any distance). Twenty executions of the model are presented. The model parameters: the size of the excised segment is 1 gene (the probability is 0.5), 2–6 (0.7), 7–14 (0.1), or 15–60 (0.05) genes. Transposition distance: 0 (the probability is 0.5), 1 (0.02), 2–6 (0.08), 7–14(0.2), or 15–60 (0.2).

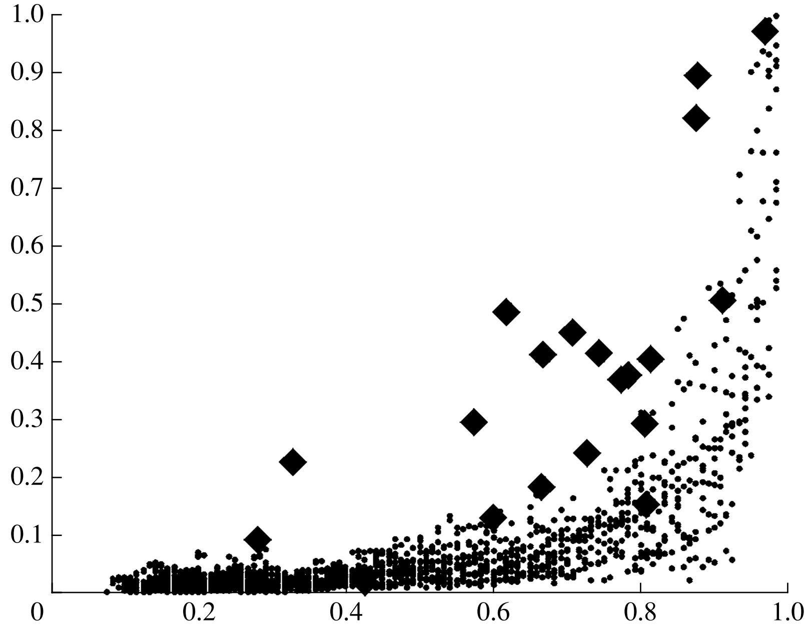

Fig. 4. Model: the rearrangement size is random (the size of the excised segment is random and the transposition distance is the same as in Fig. 3). Twenty executions of the model are presented.

ues (wMe/wBm, sus/sut, ric/rif, syn/pro, and bah/lii) fit the hypothesis of absolutely random sizes of all genome rearrangements (hypothesis 2). Furthermore, all species pairs, without any exception, satisfy the hypothesis that large rearrangements are possible but less probable compared to small rearrangements (hypothesis 3).

Thus, all genome pairs studied can be divided into three groups.

(1) Pairs with very high similarity of approximate gene orders satisfy hypotheses 1 and 3 (pyf/pyh, spn/spy, eco/shg, msa/msm, clt/cpp, and tpa/tpv). During the divergence of these pairs, the number of large rearrangements was minimal.

(2) Pairs with moderate similarity of approximate gene orders satisfy only hypothesis 3 (ana/ehr, eco/sal, eco/yep, bah/bas, cla/clp, eco/pas, myt/myl, and myt/cor).

(3) Pairs with minimal similarity of approximate gene orders satisfy hypotheses 2 and 3 (wMe/wBm, sus/sut, ric/rif, syn/pro, and bah/lii). During the divergence of these pairs, the number of large genome rearrangements was maximal.

The pair ric/rip satisfies all three hypotheses (due to the very small number of genome rearrangements having occurred in this pair). Within the limits of the gene sample examined, the differences between these two *Rickettsia* species are limited to two transpositions of short fragments (two and four genes, respectively) over a short distance.

POSSIBLE DIFFERENCES IN THE MECHANISMS AND FACTORS OF GENOME REARRANGEMENTS

If the differences in the relative frequencies of large rearrangements are, at least partly, determined not only by random events but also by differences in the mechanisms and factors of genome rearrangements, it can be expected that the groups with minimum and maximum numbers of large rearrangements strongly differ in real genetic or ecological parameters that can cause genome rearrangements.

Taxonomic position weakly correlates with the differences observed. For instance, among α-proteobacteria there are pairs with high (wMe/wBm and ric/rif) and medium (ana/ehr) numbers of genome rearrangements; among Firmicutes there are pairs with high (bah/lii), medium (bah/bas and cla/clp), and low (spn/spy) numbers of large rearrangements. On the other hand, all three pairs belonging to Euryarchaeota (pyf/pyh, msa/msm, and tpa/tpv) are characterized by a low number of large rearrangements. The only pair belonging to Crenarchaeota (sus/sut) underwent a remarkably high number of such rearrangements during its divergence.

Habitation conditions apparently have no determining influence on the relative frequency of large rearrangements, since *Sulfolobus* and *Thermoplasma*, sharing one biotope, are substantially different in the relative value of $1 - 6L/n$. Similarly strong differences were observed in intracellular parasites (wMe/wBm and ric/rif, high frequency of large rearrangements; ana/ehr, medium frequency; and clt/cpp, low frequency).

General rate of genome rearrangements seems to have no correlation with the relative frequencies of large and small rearrangements. This suggestion is supported by comparisons of the *S* values with the numbers of nucleotide substitutions in pairs of closely related prokaryotic species [14]. For instance, the ratio between the rate of genome rearrangements (change of the gene order) and the accumulation rate of nucleotide substitutions appeared to be very low in Chlamidiae and extremely high in the pair of Thermococci, pyh/pyf. At the same time, the pairs clt/cpp and pyh/pyf were very similar in the relative frequency of large genome rearrangements.

Phages and mobile elements. The *Wolbachia pipientis* strain wMel and *Rickettsia felis* are characterized by an abnormally high number of mobile genetic elements (MGEs) along with numerous signs of phage activity [37, 38]. This is also true for hyperthermophilic archaea from the genus *Sulfolobus*, characterized by an abnormally high number of transposons (in *S. solfataricus*), tandem repeat clusters [39–41], and a high viral infection rate [42]. The genome of *Listeria innocua* is overfilled with genes of a phage origin. The structure of the *Prochlorococcus* genome displays a decreased tendency towards genome rearrangements (the absence of transposases, site-specific recombinases, XerC-like integrases, etc.) [43]. At the same time, other cyanobacteria are characterized by a high number of transposases. For instance, in the *Synechococcus* strain WH 8102, signs of high phage activity along with intense horizontal gene transfer, including that realized with the help of phages, were reported [44]. Thus, all cases of minimal conservation of approximate gene orders in combination with maximum frequency of large rearrangements can be considered as evidence of the abundance of at least one member of each pair of MGEs of different nature in the genome, along with other signs of hyperactive recombination, including phage-induced recombination.

However, a quite similar pattern can be observed among the species pairs with a minimum frequency of large rearrangements. Detailed comparative analysis of the gene orders in the genomes of three species from the genus *Pyrococcus* showed the following [28]. In *P. furiosus*, the location of most small rearrangements correlates with that of 23 homologous IS-like elements associated with transposons, typical of this species and absent in other species of the genus, including *P. horikoshii.* It is suggested that *P. furiosus* was infected with these elements after its divergence from the lineage of *P. horikoshii* and *P. abyssi*. It seems likely that the predominance of small rearrangements in this pair is caused by specific features of this particular family of IS-like elements. Another characteristic feature of *Pyrococcus* is the decreased frequency of translocations crossing the replication start–end axis (the picture reciprocal to that observed in chlamydiae and mycobacteria) [28]. However, for *Sulfolobus solfataricus* (the group with the highest frequency of large rearrangements), it is suggested that the replication start–end axis is the barrier for the IS element transposition [40]. The unique feature of *P. furiosus* is that the directions of replication and transcription coincide in less than half of the genes (in other prokaryotes with complete genomes sequenced, such genes always constitute more than a half). All *Pyrococcus*, as well as *Sulfolobus*, species are characterized by the presence of long tandem repeat clusters. Distinctive signs of phage activity are practically absent in the genomes of *P. furiosus* and *P. horikoshii.* A large number of transposases is observed only in the first species (this is caused by “infection” with IS elements). It is suggested that the genome of *P. furiosus* is in the state of active rearrangement. The genome of *Thermoplasma* contains rather many transposase and, probably, no phage genes. Interestingly, *Thermoplasma* and *Sulfolobus solfataricus* inhabiting the same biotopes are characterized by intense lateral gene transfer. At least 252 genes of *Thermoplasma* are very similar to the genes of *Sulfolobus*, although these archaea belong to different large groups (Euryarchaeota and Crenarchaeota) [45]. Among the species pairs with a minimum frequency of large rearrangements, a small number or the absence of signs of phage activity was observed not only in tpa/tpv and pyf/pyh, but also in two other species pairs (msa/msm and clt/cpp). This is not true for the pairs eco/shg and spn/spy.

Thus, the groups with maximum and minimum conservation of approximate gene orders (and, correspondingly, with minimum and maximum frequencies of large rearrangements) share a common feature. Most of these pairs include by species (strains) whose genomes are in the state of active reorganization. It seems likely that phage activity in some cases can increase the probability of large genome rearrangements.

Thus, the analysis showed that the observed differences in relative frequency of large and small genome rearrangements in different prokaryotes displayed no strict correlation with the taxonomic positions of the species, abundance of transposons and repeated sequences, habitation conditions, and the general rate of genome rearrangements. These findings support the idea that the differences described can be explained by random events (it is, however, possible that phage activity can increase the probability of the appearance of large rearrangements in some cases). This conclusion is supported by computer simulation data, showing that the whole range of variation of the ratio between large and small rearrangements observed in the pairs of prokaryotic genomes studied can be reproduced using the same constant preset probability of large and small rearrangements (where large rearrangements are possible but less probable than the small ones).

ACKNOWLEDGMENTS

This work was supported by the Subprogram 2 of the Program 25 “The Origin and Evolution of Biosphere” of the Presidium of the Russian Academy of Sciences and by the Russian Science Support Foundation.

REFERENCES

1. Zakharov, I.A. and Valeev, A.K., Quantitative Analysis of the Mammalian Genome Evolution through Comparison of Genetic Maps, *Dokl. Akad. Nauk SSSR*, 1988, vol. 301, pp. 1213–1218.
2. Zakharov, I.A., Nikiforov, V.S., and Stepanyuk, E.V., Homology and Evolution of Gene Order: A Simple Method for Testing a Hypothesis on the Pattern of This Evolution, *Russ. J. Genet.*, 1996, vol. 32, no. 1, pp. 112–116.
3. Zakharov, I.A, Measurements of Similarity of Synteny Groups and an Analysis of Genome Rearrangements in the Evolution of Mammals, *Bioinformatics and Genome Research* (Proc. 3d Int. Conf.), Lim, H.A. and Cantor, C.R., Eds., Singapore: World Sci., 1995, pp. 107–113.
4. Zakharov, I.A., Nikiforov, V.S., and Stepanyuk, E.V., Homology and Evolution of Gene Order: Combinatory Measure of the Synteny Group Similarity and Simulation of the Evolutionary Process, *Genetika* (Moscow), 1992, vol. 28, no. 7, pp. 77–81.
5. Zakharov, I.A., Nikiforov, V.S., and Stepanyuk, E.V., Homology and Evolution of Gene Order: Simulation and Reconstruction of the Evolutionary Process, *Russ. J. Genet.*, 1997, vol. 33, no. 1, pp. 24–30.
6. Zakharov, I.A., Nikiforov, V.S., and Stepanyuk, E.V., Measuring of the Similarity of Homologous Gene Orders, *Genetika* (Moscow), 1991, vol. 27, no. 2, pp. 367–369.
7. Blanchette, M., Kunisawa, T., and Sankoff, D., Gene Order Breakpoint Evidence in Animal Mitochondrial Phylogeny, *J. Mol. Evol.*, 1999, vol. 49, pp. 193–203.
8. Sankoff, D., Leduc, G., Antoine, N., et al., Gene Order Comparisons for Phylogenetic Inference: Evolution of Mitochondrial Genome, *Proc. Natl. Acad. Sci. USA*, 1992, vol. 89, pp. 6575–6579.
9. Ehrlich, J., Sankoff, D., and Nadeau, J.H., Synteny Conservation and Chromosome Rearrangements during Mammalian Evolution, *Genetics*, 1997, vol. 147, no. 1, pp. 289–296.
10. Bourque, G. and Pevzner, P.A., Genome-Scale Evolution: Reconstructing Gene Orders in the Ancestral Species, *Genome Res.*, 2002, vol. 12, no. 1, pp. 26–36.
11. Rogozin, I.B., Makarova, K.S., Murvai, J., et al., Connected Gene Neighborhoods in Prokaryotic Genomes, *Nucleic Acids Res.*, 2002, vol. 30, no. 10, pp. 2212–2223.
12. Sankoff, D., Rearrangements and Chromosomal Evolution, *Curr. Opin. Genet. Dev.*, 2003, vol. 13, pp. 583–587.
13. Sankoff, D. and Nadeau, J.H., Chromosome Rearrangements in Evolution: From Gene Order to Genome Sequence and Back, *Proc. Natl. Acad. Sci. USA*, 2003, vol. 100, no. 20, pp. 11 188–11 189.
14. Suyama, M. and Bork, P., Evolution of Prokaryotic Gene Order: Genome Rearrangements in Closely Related Species, *Trends Genet.*, 2001, vol. 17, no. 1, pp. 10–13.
15. Tamames, J., Evolution of Gene Order Conservation in Prokaryotes, *Genome Biol.*, 2001, vol. 2, no. 6, pp. 0020.1–0020.11.
16. Tang, J. and Moret, B.M.E., Scaling up Accurate Phylogenetic Reconstruction from Gene-Order Data, *Bioinformatics*, 2003, vol. 19, Suppl. 1, pp. i305–i312.
17. Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E.V., Genome Alignment, Evolution of Prokaryotic Genome Organization, and Prediction of Gene Function Using Genomic Context, *Genome Res.*, 2001, vol. 11, no. 3, pp. 356–372.
18. Korbel, J.O., Snel, B., Huynen, M.A., and Bork, P., SHOT: A Web Server for the Construction of Genome Phylogenies, *Trends Genet.*, 2002, vol. 18, pp. 158–162.
19. Zakharov, I.A. and Markov, A.V., Gene Orders in Genomes of α-Proteobacteria: Similarity and Evolution, *Russ. J. Genet.*, 2005, vol. 41, no. 12, pp. 1343–1351.
20. Alm, R.A., Ling, L.S., Moir, D.T., et al., Genomic-Sequence Comparison of Two Unrelated Isolates of the Human Gastric Pathogen *Helicobacter pylori, Nature*, 1999, vol. 397, no. 6715, pp. 176–180.
21. Eisen, J.A., Heidelberg, J.F., White, O., and Salzberg, S.L., Evidence for Symmetric Chromosomal Inversions Around the Replication Origin in Bacteria, *Genome Biol.*, 2000, vol. 1, no. 6, p. RESEARCH0011.
22. Grigoriev, A., Graphical Genome Comparison: Rearrangements and Replication Origin of *Helicobacter pylori, Trends Genet.*, 2000, vol. 16, no. 9, pp. 376–378.
23. Hughes, D., Evaluating Genome Dynamics: The Constraints on Rearrangements within Bacterial Genomes, *Genome Biol.*, 2000, vol. 1, no. 6, p. REVIEWS0006.
24. Maeder, D.L., Weiss, R.B., Dunn, D.M., et al., Divergence of the Hyperthermophilic Archaea *Pyrococcus furiosus* and *P. horikoshii* Inferred from Complete Genomic Sequences, *Genetics*, 1999, vol. 152, no. 4, pp. 1299–1305.
25. Read, T.D., Brunham, R.C., Shen, C., et al., Genome Sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39, *Nucleic Acids Res.*, 2000, vol. 28, no. 6, pp. 1397–1406.
26. Tillier, E.R. and Collins, R.A., Replication Orientation Affects the Rate and Direction of Bacterial Gene Evolution, *J. Mol. Evol.*, 2000, vol. 51, no. 5, pp. 459–463.
27. Tillier, E.R. and Collins, R.A., Genome Rearrangement by Replication-Directed Translocation, *Nat. Genet.*, 2000, vol. 26, no. 2, pp. 195–197.
28. Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P., *Pyrococcus* Genome Comparison Evidences Chromosome Shuffling-Driven Evolution, *Nucleic Acids Res.*, 2002, vol. 30, no. 9, pp. 1902–1910.
29. Huynen, M.A., Snel, B., and Bork, P., Inversions and the Dynamics of Eukaryotic Gene Order, *Trends Genet.,* 2001, vol. 17, no. 6, pp. 304–306.
30. Dalevi, D.A., Eriksen, N., Eriksson, K., and Andersson, S.G., Measuring Genome Divergence in Bacteria: A Case Study Using Chlamydian Data, *J. Mol. Evol.*, 2002, vol. 55, no. 1, pp. 24–36.
31. Sankoff, D., Short Inversions and Conserved Gene Clusters, *Bioinformatics*, 2002, vol. 18, no. 10, pp. 1305–1308.

32. Durand, D. and Sankoff, D., Tests for Gene Clustering, *J. Comput. Biol.*, 2003, vol. 10, nos. 3–4, pp. 453–482.

33. Trachtulec, Z. and Forejt, J., Synteny of Orthologous Genes Conserved in Mammals, Snake, Fly, Nematode, and Fission Yeast, *Mamm. Genome*, 2001, vol. 3, no. 12, pp. 227–231.

34. Brinkman, F.S.L., Blanchard, J.L., Cherkasov, A., et al., Evidence that Plant-Like Genes in *Chlamydia* Species Reflect an Ancestral Relationship between Chlamydiaceae, Cyanobacteria, and the Chloroplast, *Genome Res.*, 2002, vol. 12, pp. 1159–1167.

35. Coenye, T. and Vandamme, P., Organisation of the S10, *spc* and α Ribosomal Protein Gene Clusters in Prokaryotic Genomes, *FEMS Microbiol. Lett.*, 2005, vol. 242, pp. 117–126.

36. Hauth, A.M., Maier, U.G., Lang, B.F., and Burger, G., The *Rhodomonas salina* Mitochondrial Genome: Bacteria-Like Operons, Compact Gene Arrangement and Complex Repeat Region, *Nucleic Acids Res.*, 2005, vol. 33, no. 14, pp. 4433–4442.

37. Ogata, H., Renesto, P., Audic, S., et al., The Genome Sequence of *Rickettsia felis* Identifies the First Putative Conjugative Plasmid in an Obligate Intracellular Parasite, *PLoS Biol.*, 2005, vol. 3, no. 8, p. e248.

38. Wu, M., Sun, L.V., Vamathevan, J., et al., Phylogenomics of the Reproductive Parasite *Wolbachia pipientis* wMel: A Streamlined Genome Overrun by Mobile Genetic Elements, *PLoS Biol.*, 2004, vol. 2, no. 3, pp. 327–341.

39. Blount, Z.D. and Grogan, D.W., New Insertion Sequences of *Sulfolobus*: Functional Properties and Implications for Genome Evolution in Hyperthermophilic Archaea, *Mol. Microbiol.*, 2005, vol. 55, no. 1, pp. 312–325.

40. Brugger, K., Redder, P., She, Q., et al., Mobile Elements in Archaeal Genomes, *FEMS Microbiol Lett.*, 2002, vol. 206, no. 2, pp. 131–141.

41. She, Q., Singh, R.K., Confalonieri, F., et al., The Complete Genome of the Crenarchaeon *Sulfolobus solfataricus* P2, *Proc. Natl. Acad. Sci. USA,* 2001, vol. 98, no. 14, pp. 7835–7840.

42. Wiedenheft, B., Stedman, K., Roberto, F., et al., Comparative Genomic Analysis of Hyperthermophilic Archaeal Fuselloviridae Viruses, *J. Virol.*, 2004, vol. 78, no. 4, pp. 1954–1961.

43. Dufresne, A., Salanoubat, M., Partensky, F., et al., Genome Sequence of the Cyanobacterium *Prochlorococcus marinus* SS120, a Nearly Minimal Oxyphototrophic Genome, *Proc. Natl. Acad. Sci. USA*, 2003, vol. 100, no. 17, pp. 10 020–10 025.

44. Palenik, B., Brahamsha, B., Larimer, F.W., et al., The Genome of a Motile Marine *Synechococcus, Nature*, 2003, vol. 424, no. 6952, pp. 1037–1042.

45. Ruepp, A., Graml, W., Santos-Martinez, M.L., et al., The Genome Sequence of the Thermoacidophilic Scavenger *Thermoplasma acidophilum, Nature*, 2000, vol. 407, no. 6803, pp. 508–513.