Biomedical Sciences Today
An open access peer reviewed journal
MDT Canada press
http://www.mdtcanada.ca/bmst.html

Regular Article

Analyzing Inverse Symmetry with Original and Terminal Sites of Prokaryotic Genome

Tzu-Ting Hsu[1,*], Hoong-Chien Lee[1]
[1]Graduate Institute of Systems Biology and Bioinformatics, National Central University, Chungli, Taiwan, Republic of China
*Correspondence Email: lecia.hsu@gmail.com

Received 12 November 2015; Revised 1 December 2015; Accepted 1 December 2015; Published 25 December 2015

Editor: Mohammad Ashrafuzzaman

Abstract

The segmental duplication has long been known to be a major mechanism for genome growth and evolution. Beyond that, the large-scale genomic duplication is important in the evolution of species. Recent studies have spotted occurrence of whole-genome duplication provided great insight into many aspects of biology. In this work we investigate possible association of inverse symmetry with whole-genome inverse duplications on genomes. Our findings, through analyses of word-frequency nucleotides and distributions of homologous conserved regions on publicly available complete genomes of 18 archaea and 139 bacteria, are positive. These findings suggest that, first, statistically significant inverse symmetry is found within some prokaryotic genomes. Second, there is inverse symmetry around the replication original and terminal sites on the prokaryotic genomes. The mechanism of generation of inverse symmetry involves whole-genome inverse duplications. Whole-genome inverse duplications have occurred in most prokaryotic chromosomes near original sites and terminal sites. Our research integrates the abundance of knowledge with evolution of genomes and creates a new approach for predicting loci of original sites and terminal sites in prokaryotic chromosomes. Whole-genome inverse duplication events probably are ancient, may have occurred not frequently but probably in more than one incident. Our findings suggest that whole-genome duplication is a great feature of prokaryotic genome evolution.

Keywords: Replication, Prokaryotic Genome, Inverse Symmetry, Terminal Site, Original Site

Biomedical Sciences Today
An open access peer reviewed journal
MDT Canada press
http://www.mdtcanada.ca/bmst.html

1. Introduction

Chargaff's first parity rule, stating that in a DNA molecule contents of adenine (A) and thymidine (T), and of cytosine (C) and guanine (G), are separately identical, was proposed in 1953 [1,2]. It was not only a crucial clue to Watson and Crick's discovery of the double helical structure of DNA [3], but also a basis for molecular biology. In 1968, Chargaff proposed Chargaff's second parity rule which states that at a lower level of accuracy the first rule also extends to a single strand of DNA [4]. Chargaff's parity rule is a case of special symmetry. Symmetry exists in mathematics [5], physics [6], chemistry [7], evolution [8], human appearance [9], and psychology [10]. We consider the 3-mer ATC. Its reverse (r), complement (c), and inverse (i) conjugates are CTA, TAG, and GAT, respectively. According to the current research, Chargaff's second parity rule is a special case of inverse symmetry, not complement symmetry [11-14].

Recently we conducted a comprehensive study of symmetry, including inverse symmetry, in complete genomes [15]. In this study, we find that as a rule neither reverse nor complement symmetry exist in genomes, but a vast majority of genomes have close to maximum global inverse symmetry, while exhibiting starkly distinct patterns of local inverse symmetry. These genomes possess abundant inverse symmetry, but not contain reverse symmetry and complement symmetry. It suggests Chargaff's second parity rule is a special case of inverse symmetry. According to the current understanding, the segmental duplication is the force in chromosome growth and evolution [16-21], and inverse segmental duplication events have occurred in chromosome evolution [22-24]. We believe that the mechanism of the inverse duplication contributes to inverse symmetry [15].

2. Materials and Methods

We call a k-base nucleic word a k-mer. Given a sequence, we count the frequency of occurrence fu of each k-mer type u using an overlapping sliding window of width k and slide one [42]. Given k, let P_ρ be the set of distinct ρ-conjugate but non-self conjugate pairs of k-mers types, where ρ=r, c, and i denote reverse, complement, and inverse symmetry, respectively. The ρ-symmetry index, χ_ρ, is defined as:

$$\chi_\rho^2 = \frac{1}{2N_\rho} \sum_{(u,u^+)\in P_\rho} \left(\frac{f_u - f_{u^+}}{\sigma_{m_u}} \right)^2 \ldots \rho = \text{r, c, or i}$$

Where u+ is the ρ-conjugation of u, σ_{mu} is the standard deviation of the frequencies of k-mers in the m-set to which both u and u+ belong, and N_ρ is the number of ρ conjugate pairs in P_ρ. By design χ_ρ is expected to be close to unity in the absence of ρ symmetry. A χ_ρ significantly less than unity indicates the presence of ρ-symmetry and $\chi_\rho = 0$ implies exact ρ-symmetry.

3. Results

We intend to include in this study all the non-redundant prokaryotic complete sequences in public databases. The 157 complete sequences of 28 archaea chromosomes and 139 eubacteria chromosomes have been analyzed in this study. Those are downloaded from the National Center for Biotechnology Information (NCBI) chromosome database [25]. Individual chromosomes range in length from 490 kb to 9105 kb.

The quantity χ_ρ measures the segmental average of the ρ-symmetry index for a chromosome partitioned into segments of length (Methods). The smaller χ_ρ means there is a better ρ-symmetry on the sequence. We compute the χ_ρ for 157 prokaryotic chromosomes downloaded from NCBI in Table S1, SI. Here we present the evidence for the ρ-symmetry index for four species of Clostridium acetobutylicum, Erwinia carotovora, Escherichia coli, and Synechocystis in Table 1. The Table 1 is not only a summary of the finding that reverse and complement symmetries are absent, but also a result showing that the inverse symmetry is strongly present.

Table 1. There is neither reverse nor complement symmetry in C. acetobutylicum, E. carotovora, E. coli, and Synechocystis. The perfect inverse symmetry in these four species is shown in the table.

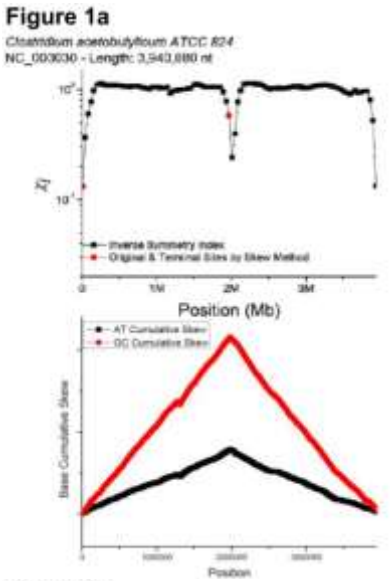
Table 1

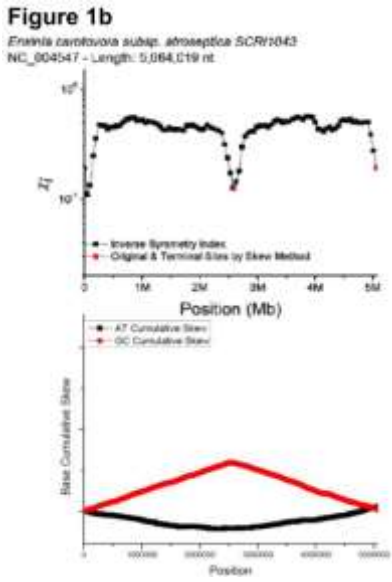
Organism Name	Length	χ_r	χ_c	χ_i
C. acetobutylicum	3940880	1.0415	0.9025	0.0387
E. carotovora	5064019	1.1609	1.0056	0.0369
E. coli	4639675	1.1091	0.9607	0.0255
Synechocystis	3573470	0.8546	0.7401	0.0263

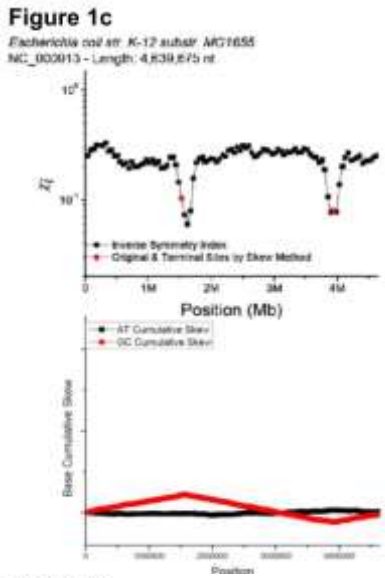
The χ_i scanning plots for complete genomes of C. acetobutylicum, E. carotovora, E. coli, and Synechocystis are shown in Figure 1. There are

Biomedical Sciences Today
An open access peer reviewed journal
MDT Canada press
http://www.mdtcanada.ca/bmst.html

isolated sharp minima, prominent in C. acetobutylicum and E. carotovora and more numerous or less conspicuous in E. coli. but absent in Synechocystis. We believe that the weak inverse symmetry could be explained by gene loss that has been found on chromosomal or genome duplications [26-28]. Recently replication original and terminal sites have been found from analysis of GC cumulative skew method [29]. We find that original and terminal sites, already identified as break-points in inverse symmetry, are near centers of inverse symmetry reflection in Figure 1. Therefore, we could take χ_i as a feature to predict original and terminal sites of replication and illustrate the results.







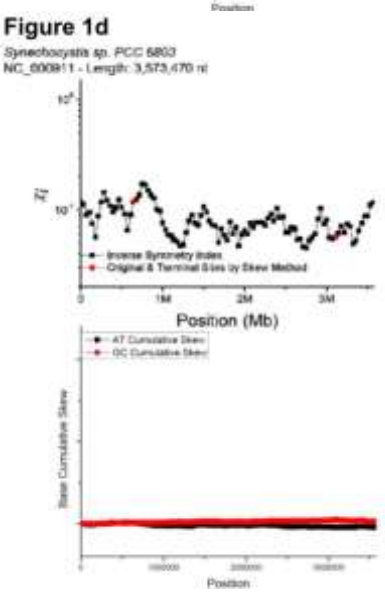


Figure 1. The χ_i scanning and the skew method plots for four representative organisms, C. acetobutylicum, E. carotovora, E. coli, and Synechocystis of Figure 1a, 1b, 1c, and 1d respectively. The upper part presents the whole-genome χ_i scanning. The lower part presents the AT and GC cumulative skew on the genome. The GC cumulative skew is used to predict the replication original and terminal sites. The perfect inverse symmetry are found around original and terminal sites in C. acetobutylicum, E. carotovora, and E. coli. The weak or miss inverse symmetry near original and terminal sites in Synechocystis is also indistinct by using GC cumulative skew.

4. Discussion

Because inverse symmetry is decided by all the letters in a k-mer as a group, its prevalence cannot

Biomedical Sciences Today
An open access peer reviewed journal
MDT Canada press
http://www.mdtcanada.ca/bmst.html

be solely the result of point mutations or any mechanism that affects one nucleotide at a time. Ubiquitous genomic inverse symmetry must be the manifest of a certain mode of genome growth and evolution. We have proposed a hypothesis to explain why inverse symmetry exists in species. Consider a segment S in the positive strand of the chromosome and its complement S_c in the negative strand. Now suppose S_c is copied and inserted somewhere into the positive strand. Because of the backbone structure of a DNA strand, S_c must be reversed before the insertion. That is, the inverse of S, namely S_i, is actually inserted in Figure 2. Inverse duplication generates inverse symmetry. We believe inverse symmetry variations in Figure 1 are due to different numbers of inverse duplication events occurred in species. No other comparable mechanism exists that can generate either reverse or complement symmetry, and this may explain why they are not observed in genomes.

Figure 2

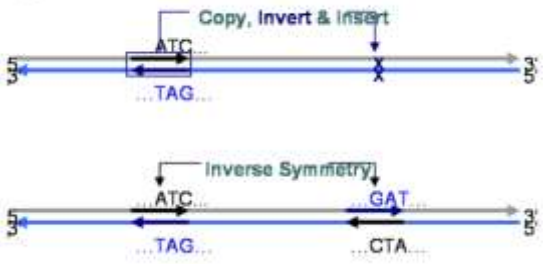


Figure 2. A double-strands DNA sequence includes a positive strand from 5' to 3' and a negative strand from 3' to 5'. They are complement conjugation. We assume S, S_c, and S_i as ATC, TAG, and GAT segment, respectively.

The whole-genome duplication is an important mechanism for genome growth and evolution. It has been studied through comparative genetics research. Occurrence of the whole-genome duplication was suggested by analysis of yeast [30–32], ray-finned fishes [33], and freshwater puffer fish [34]. The whole-genome inverse duplication was also been suggested for E. coli [35]. Recently it has been found from analysis of B. burgdorferi [36,37] that led to much better insight into the whole-genome inverse duplication. In our study, we find that the inverse symmetry in species was consistent with occurrence of the whole-genome inverse duplication that pivoted around original and terminal sites in Figure 1. This finding suggest that an inverse symmetry within species is the ancestral inverse duplication of the whole genome occurred near the original and terminal site. Inverse symmetry is also a new feature to predict loci of original and terminal sites.

If the whole-genome inverse duplication has occurred around original and terminal sites, we could observe the positive strand sequence from the original site to the terminal site is similar to the negative strand sequence from the original site to the terminal site. To compare the conversed regions of C. acetobutylicum, we divide the complete sequence into two halves. The upper half is the sequence from the original site to the terminal site and the lower half is the sequence from the terminal site to the origin site. We analyze those two sequences of C. acetobutylicum by using Mauve [38]. Mauve is useful for browsing large scale genome with genomic rearrangements on the level of DNA. A whole-genome comparison of genome positions of conversed regions within C. acetobutylicum is shown in Figure 3a. Blocks above the center line mean the aligned regions are the upper half sequence in the positive strand. Blocks below the center line indicate the aligned regions are in the negative strand. The alignment of C. acetobutylicum consists of 10 locally collinear conversed blocks that cover an average of 91.7% of each genome as Figure 3a. Figure 3a also reveals a large-scale inverse symmetry between the upper and lower half sequences. In addition, the inverse symmetry is not limited to the C. acetobutylicum comparison. Inverse symmetry has also been found in Clostridium novyi in Figure 3b. These results present evidence that the whole-genome inverse duplication event has occurred around original and terminal sites.

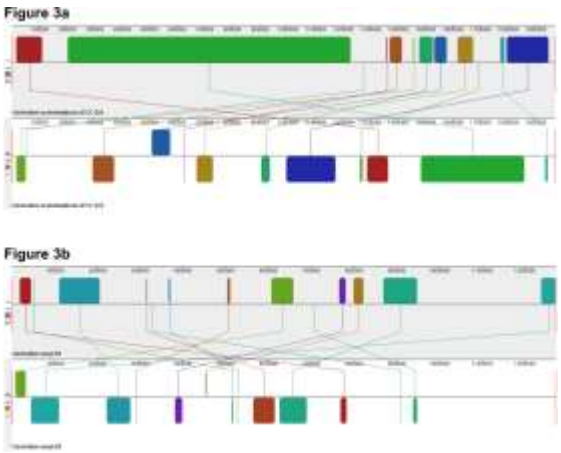


Figure 3. Conversed regions of C. acetobutylicum and C. novyi are prsesnted in Figure 3a and 3b. In both Figure 3a and 3b, the upper direct sequence is similar to the lower indirect sequence. These results show whole-genome inverse duplication events have existed in genome evolution.

Biomedical Sciences Today
An open access peer reviewed journal
MDT Canada press
http://www.mdtcanada.ca/bmst.html

Whole-Genome alignments could align genome sequences with duplications and rearrangements. To observe the duplication and rearrangement events of genome growth and evolution, we align the sequence of C. acetobutylicum with itself by zPicture [39], a sequence analysis tool using BLASTZ [40]. After generation of local alignments between the two sequences, we construct of a rough global map by chaining an ordered subset of the local alignments. The analysis reveals a significant 'X' shape alignment Figure 4a. If the whole-genome inverse duplication is a real evolutionary events in the lineage, we could expect not only the regions along one diagonal but also the regions along the other diagonal should be perfect homologous within species. In Figure 4a, there is a significant X-alignment within the genome of C. acetobutylicum. The X-alignment within the species of C. novyi is also shown that there is an ancestral inverse duplication around original and terminal site in Figure 4b. We believe that the weak anti-diagonal in Figure 4a and 4b are due to many rearrangement events in the different lineages [41] in Figure 4c.

Figure 4a

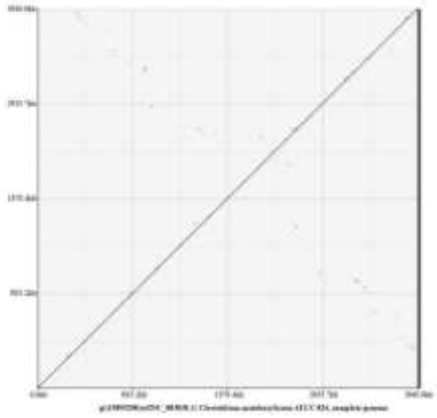


Figure 4b

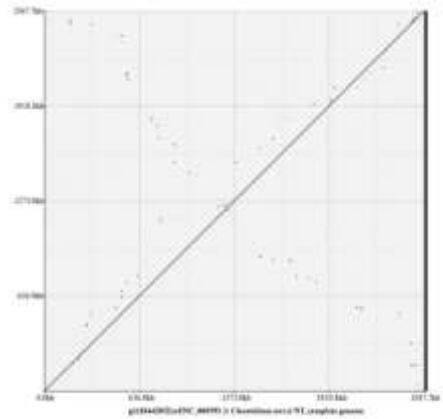


Figure 4c

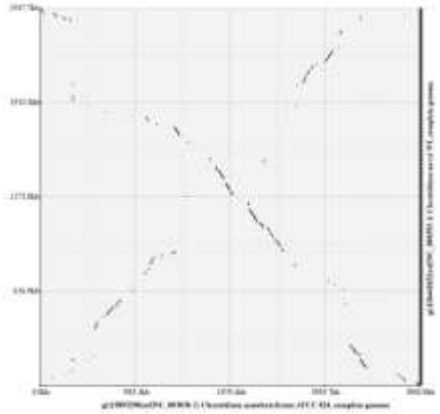


Figure 4. In the dot-plots, the reference sequence is laid across the X-axis, while the query sequence is on the Y-axis, and a point is plotted at every position where the two sequences show similarity. If the two sequences are perfectly identical, a diagonal would go from the bottom left to the top right.

We think that original and terminal sites, already identified as break- points in inverse symmetry, are near centers of inverse-symmetry reflection. The correlation between type-classification and phylogeny is illustrated in Fig. 2 by using iTOL [43, 44]. The colors of bars, blue, green, yellow, and red, are for types A, B, C, and D, respectively. There is a dominating type in each monophyletic category.

5. Conclusion

Ancient events in evolution intend to be obscured or masked by intervening mutation events, thus making analysis difficult. We present a novel observation regarding the genome growth and evolution on prokaryotic species. Our studies reveal that whole genome inverse duplication events are occurred around original and terminal site. The difference of inverse symmetry between species could be the different numbers of inverse duplication events. These results also suggest that the whole-genome inverse duplication is a common feature of prokaryotic evolution. We could also take the inverse symmetry index as a feature to predict the original and terminal site of replication in prokaryotic genomes. A public online tool for computing the inverse symmetry index and predicting the original and terminal sites of prokaryotic organisms will be developed. We will conduct a detailed study of the correlation between type-classification and phylogeny. We could estimate the time when whole genome inverse

duplication occurred by combining result on whole-genome inverse duplication with phylogenetic information. Infer from the study a narrative of genome growth and evolution, and a biological interpretation of this narrative.

Conflict of Interests

The authors declare no conflict of interests regarding the publication of this article.

Acknowledment

This work is supported in part by grant nos. 96-2112-M-008-025 and 97-2112-M-008013, National Science Council, ROC (http://web1.nsc.gov.tw/), and the Cathy General Hospital-National Central University (http://www.ncu.edu.tw/e_web/index.php) Grant 96-CGH-NCU-A1.

References

[1] Chargaff E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia 1950, 6:201-209.
[2] Chargaff E. Structure and function of nucleic acids as cell constituents. Fed Proc 1951, 10:654-659.
[3] Watson JD, Crick FH. Genetical implications of the structure of deoxyribonucleic acid. Nature 1953, 171:964-967.
[4] Rudner R, Karkas JD, Chargaff E. Separation of B. subtilis DNA into complementary strands, I. Biological properties. Proc. Natl. Acad. Sci. U S A 1968, 60:630-635.
[5] Dixon L, Ginsparg P, Harvey J. Beauty and the beast: Superconformal symmetry in a monster module. Comm Math Phys 1988, 119: 221-241.
[6] Zee A. Fearful Symmetry: The Search for Beauty in Modern Physics. Princeton: University Press 1999, 336 p.
[7] Mueller A. Chemistry. The beauty of symmetry. Science 2003, 300: 749750.
[8] Enquist M, Arak A. Symmetry, beauty and evolution. Nature 1994, 372: 169-172.
[9] Grammer K, Thornhill R. Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. J Comp Psychol 1994, 108: 233-242.
[10] Rhodes G. The evolutionary psychology of facial beauty. Annu Rev Psychol 2006, 57: 199-226.
[11] Qi D, Cuticchia AJ. Compositional symmetries in complete genomes. Bioinformatics 2001, 17: 557-559.
[12] Forsdyke DR. Symmetry observations in long nucleotide sequences: a commentary on the Discovery Note of Qi and Cuticchia. Bioinformatics 2002, 18: 215-217.
[13] Baisnee PF, Hampson S, Baldi P. Why are complementary DNA strands symmetric? Bioinformatics 2002, 18: 1021-1033.
[14] Jose MV, Govezensky T, Bobadilla JR. Statistical properties of DNA sequences revisited: the role of inverse bilateral symmetry in bacterial chromosomes. Physica 2005, 351: 477-498.
[15] Kong SG, Fan WL, Chen HD, Hsu TT, Zhou N, Zheng B, Lee HC. Inverse symmetry in genomes and whole-genome inverse duplication. PLoS ONE 2009, 4: e7553.
[16] Lynch M. Gene duplication and evolution. Science 2002, 297: 945–947.
[17] Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. Recent segmental duplications in the human genome. Science 2002, 297: 1003–1007.
[18] Hsieh LC, Luo L, Ji F, Lee HC. Minimal model for genome evolution and growth. Phys Rev Lett 2003, 90: 018101.
[19] Zhang J. Evolution by gene duplication: an update. Trends Eco Evolut 2003, 18: 292–298.
[20] Zhang L, Lu HH, Chung WY, Yang J, Li WH. Patterns of segmental duplication in the human genome. Mol Biol Evol 2005, 22: 135–141.
[21] Messer PW, Arndt PF, Lassig M. Solvable sequence evolution models and genomic correlations. Phys Rev Lett 2005, 94: 138103.
[22] Nussinov R. Some indications for inverse DNA duplication. J Theor Biol 1982, 95: 783–791.
[23] Biebricher CK, Luce R. In vitro recombination and terminal elongation of RNA by Q beta replicase. EMBO J 1992, 11: 5129–5135.
[24] Volz A, Wende H, Laun K, Ziegler A. Genesis of the ILT/LIR/MIR clusters within the human leukocyte receptor complex. Immunol Rev 2001, 181: 39–51.
[25] National center for biotechnology information genome database. Available: http://www.ncbi.nlm.nih.gov/.
[26] Wagner A. The fate of duplicated genes: loss or new function? Bioessays 1998, 20:785-788.
[27] Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. Genetics 2000, 154:459-473.
[28] Nadeau JH, Sankoff D. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 1997, 147:1259-1266
[29] Grigoriev A. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 1998, 26:2286-2290.

[30] Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 1997, 387: 708–713.
[31] Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 2004, 428: 617–624.
[32] Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. Nature 2007, 449: 54–61.
[33] Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol 2004, 21: 1146–1151.
[34] Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, RobinsonRechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate protokaryotype. Nature 2004, 431: 946–957.
[35] Zipkas D, Riley M. Proposal concerning mechanism of evolution of the genome of Escherichia coli. Proc Natl Acad Sci U S A 1975, 72:1354-1358.
[36] Jose MV, Govezensky T, Bobadilla JR: Statistical properties of DNA sequences revisited. the role of inverse bilateral symmetry in bacterial chromosomes. Physica 2005, 351: 477–498.
[37] Sanchez J, Jose MV. Analysis of bilateral inverse symmetry in whole bacterial chromosomes. Biochem Biophys Res Commun 2002, 99: 126–134.
[38] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004, 14:1394-1403.
[39] Ovcharenko I, Loots GG, Hardison RC, Miller W, Stubbs L: zPicture. Dynamic alignment and visualization tool for analyzing conservation profiles. Genome Res 2004, 14:472-477
[40] Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. Human–mouse alignments with BLASTZ. Genome Res 2003, 13: 103–107.
[41] Eisen JA, Heidelberg JF, White O, Salzberg SL. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol 2000, 1:RESEARCH0011.
[42] Hao BL, Lee HC, Zhang SY. Fractals related to long DNA sequences and complete genomes. Chaos, Solitons & Fractals 2000, 11:825-836.