Mol Biol Rep (2008) 35:163–169
DOI 10.1007/s11033-007-9066-6

Strand compositional asymmetries in vertebrate large genes

Hai-Fang Wang · Wen-Ru Hou · Deng-Ke Niu

Received: 8 November 2006 / Accepted: 26 February 2007 / Published online: 10 April 2007

Abstract Both transcription-associated and replication-associated strand compositional asymmetries have recently been shown in vertebrate genomes. In this paper, we illustrate that transcription-associated strand compositional asymmetries and replication-associated ones coexist in most vertebrate large genes, although in most case the former conceals the latter. Furthermore, we found that the transcription-associated strand compositional asymmetries of housekeeping genes are stronger than those of somatic cell expressed genes. Together with other evidence, we suggest that germline transcription-associated strand asymmetric mutations may be the main cause of the transcription-associated strand compositional asymmetries.

Keywords Cumulative skew diagram · Illegitimate transcription · Replication · Transcription

Abbreviations

CSD	Cumulative skew diagram
CSDD	Cumulative skew difference diagram

Electronic Supplementary Material The online version of this article (doi:10.1007/s11033-007-9066-6) contains supplementary material, which is available to authorized users.

H.-F. Wang · W.-R. Hou · D.-K. Niu (✉)
Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China
e-mail: dkniu@bnu.edu.cn

Present Address:
W.-R. Hou
National Institute of Biological Sciences, Beijing 102206, China

Introduction

In prokaryotic, organelle, and viral genomes, the leading strand was found to have more G than C and to a lesser extent more T than A [1–9]. These compositional biases were mainly attributed to asymmetric nucleotide substitution patterns associated with the structural asymmetries of DNA replication processes [6, 10, 11]. Transcription increases the rate of deamination on the coding strand and favors transcription-associated repair on the template strand [2, 12, 13], thus would also produce strand compositional asymmetries. It was observed that the majority of bacterial genes (especially for essential genes) are transcribed co-directionally with replication [14–16]. Baran and Ko [17] have presented a mathematical model to quantify the preference of genes to be encoded on the leading strand. Recently, Nikolaou and Almirantis [18] found a close correlation between the strand nucleotide asymmetry and the asymmetric orientation of coding sequences throughout the bacterial genomes. It seems that transcription-associated mutational and/or selective pressures, rather than the replication-associated ones, may be the main source of strand compositional asymmetries found in prokaryotic genomes. Whereas, using statistical methods to compare these effects in bacterial genomes, Baran and Ko [19] show that neither replication-coupled processes nor transcription-coupled ones can solely explain the strand compositional asymmetry in general.

In eukaryotes, indisputable results on strand compositional asymmetries were first obtained in studying the transcribed regions of the genomes [20, 21]. Further evidence suggests that transcription-associated strand asymmetries in eukaryotes are generally much stronger than the replication-associated asymmetries if they exist [22]. Replication-associated strand compositional asymmetries

in vertebrate genomes were recently revealed by analyzing the intergenic sequences [23–25]. This paper is to illustrate that both transcription-associated and replication-associated strand compositional asymmetries exist in most transcribed regions of vertebrate genomes, although in most case the former conceals the latter.

Transcription-associated strand compositional asymmetry may result from strand asymmetric mutational pressure (e.g. transcription-coupled repair [20] and transcription-coupled mutagenesis [26]) and/or strand asymmetric selective pressures (e.g. codon bias or maintenance of functional elements for splicing or regulating gene expression) [6, 27–30]. Obviously, if mutational pressures are the main cause of transcription-associated strand compositional asymmetry, one would expect a difference in strand compositional asymmetry between genes expressed in germline cells and those expressed only in somatic cells. Such evidence will be reported in this paper.

Materials and methods

All the annotated eukaryotic genomes were retrieved from the NCBI GenBank database (ftp://ftp.ncbi.nih.gov). As discussed previously [25], only large genes that are expected to span at least one replicon are appropriate to reveal replication-associated asymmetries by common methods, like cumulative skew diagram (CSD) [3], DNA walk [31], and *Z* curve [32]. Eukaryotic replicons are generally believed to be 40–100 kb in length [25, 33], so only genes larger than 50 kb were parsed out. We found sufficient number of genes larger than 50 kb only in vertebrate genomes, *Danio rerio* (NCBI build 1 version 1), *Gallus gallus* (NCBI build 1 version 1), *Mus musculus* (NCBI build 32), *Rattus norvegicus* (NCBI build 2), *Canis familiaris* (NCBI build 1 version 1), *Pan troglodytes* (NCBI build 1 version 1), and *Homo sapiens* (NCBI build 34 version 3). Genes with alternative splicing sites, with obvious annotation errors, or that overlapped other genes were excluded from our analysis.

In large genes, transcription-associated strand compositional asymmetries are expected to be constant along the gene while replication-associated ones should switch signs at replication origins and termini. The combined strand compositional asymmetries were shown by CSD [3] with small modifications as described in our previous paper [25]. Previously, we estimated replicon sizes by the positions of distinct global extrema in the cumulative AT skew diagrams of large intergenic sequences [25]. There are two reasons for us to use IGSs' CSD to predict the size of replicons. First, intergenic sequences are not expected to be transcribed, and thus their asymmetries would not be associated with transcription. Second, CSD is the most

Table 1 Percentage of vertebrate large genes with different types of cumulative skew diagrams (CSDs)

Species	Number of genes studied	AT Skew					CG Skew				
		V-shaped CSDs	Disordered CSDs	/-shaped CSDs	\-shaped CSDs	V-shaped CSDDs[a]	V-shaped CSDs	Disordered CSDs	/-shaped CSDs	\-shaped CSDs	V-shaped CSDDs[a]
Danio rerio	1227	0.335	0.093	0.044	0.528	0.835	0.459	0.139	0.032	0.370	0.830
Gallus gallus	1205	0.094	0.022	0.032	0.853	0.891	0.175	0.022	0.026	0.778	0.905
Canis familiaris	1422	0.200	0.036	0.052	0.712	0.972	0.233	0.054	0.053	0.659	0.939
Mus musculus	1543	0.194	0.192	0.016	0.598	0.883	0.366	0.201	0.021	0.412	0.684
Rattus norvegicus	1598	0.210	0.227	0.031	0.533	0.959	0.387	0.268	0.036	0.309	0.844
Pan troglodytes	1982	0.244	0.075	0.061	0.620	0.889	0.188	0.133	0.058	0.622	0.774
Homo sapiens	1684	0.219	0.112	0.017	0.651	0.966	0.287	0.068	0.014	0.631	0.869

[a] Only genes with \-shaped CSDs were counted; CSDD, cumulative skew difference diagrams

simple and intuitive method to explore the DNA strand compositional asymmetries. The skew values [$(C - G)/(C + G)$ and $(A - T)/(A + T)$] of adjacent 100 bp windows (1 kb windows give similar results) along the DNA sequence were consecutively added together and plotted. If a gene happens to cover complete replicon units, the sum of replication-associated compositional asymmetries is expected to be zero, and then the average skew value ($\bar{S}$) just reflects the strand compositional asymmetry contributed by transcription-associated processes. The strength of the transcription-associated strand compositional asymmetry can thus be roughly measured by the average value of the skews of each gene. Meanwhile, the strand compositional asymmetry contributed by replication-associated processes should be the skew difference value ($\Delta S = S - \bar{S}$). To illustrate the replication-associated strand compositional asymmetries in gene sequences, we designed a new method, the cumulative skew difference diagram (CSDD), to filter out the potential effects of transcription. In a CSDD, skew difference values (ΔS) were consecutively added and plotted. In most cases, genes are smaller or larger than complete replicon units, so replication-associated asymmetry also contributes some extent to the averaged skew value. Although that could make the CSDD profile's extrema deviated from the replication origins and termini, it will not change its global shape. Given that we just intuitively compare the effects of transcription and replication, rather than estimate the replicon sizes [25], these small deviation will have no effect on our conclusion. Only genes that have positively correlated ($P < 0.05$) AT skew and CG skew in both CSD and CSDD are used for study. As declared in our previous report [25], we identified the shaped of CSDs and CSDDs by eye. Whilst there may be some inaccuracy or error, we do not believe that these would weaken the general conclusion.

We used Affymetrix microarray data (GNF GeneAtlas version 2) [34]) to determine whether a gene is germline-cell-expressed or somatic-cell-expressed in human and mouse. As recommended [35], a gene was classified as being expressed in a tissue if its average difference value was greater than the threshold of 200 in that tissue. Housekeeping genes (that have average difference values greater than 200 in all normal tissue) were used to represent the genes that are transcribed in the germline cell. Somatic-cell-expressed genes are those expressed in less than 20% normal tissue/organ samples, but not expressed in germ line cells (oocytes or fertilized egg), reproductive organs (testis or ovary), or early developmental stages (blastocysts or embryo).

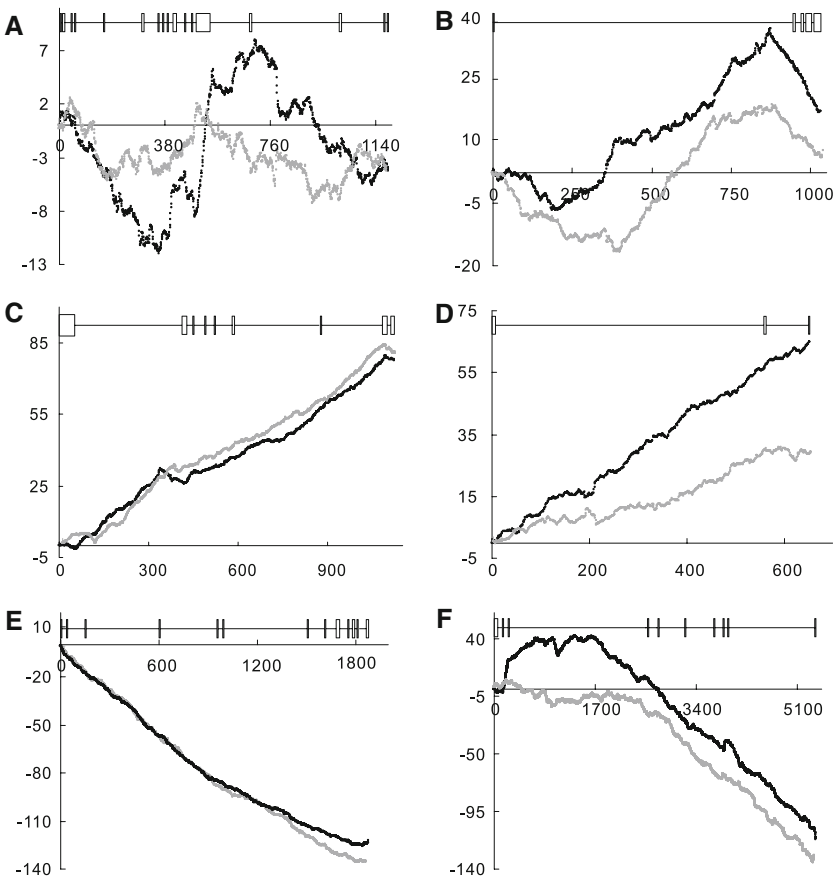

Fig. 1 Cumulative skew diagrams of some examples of vertebrate large genes. The *X* axis represents the sequence length in 100 bp units, and the *Y* axis shows the cumulative skew values. The AT skews are *black curves* and the CG skews are *grey curves*. The exon–intron structure of each gene is marked on the *top* of the figure. Exons are shown by *boxes* or *vertical bars* (in the cases where exon is very small compared with the whole gene), while introns are shown by *horizontal lines*. The sizes of boxes and lines are proportional to the length of exon sizes and intron sizes. **a** *Danio rerio* gene *LOC571429*; **b** *Mus musculus* gene *Tegt*; **c** *Gallus gallus* gene *LOC415904*; **d** *Rattus norvegicus* gene *LOC362340*; **e** *Gallus gallus* gene *LOC395163*; **f** *Pan troglodytes* gene *LOC462229*

Results and discussion

Strand compositional asymmetries of large genes

We first surveyed the proportion of genes with different types of CSDs (Table 1). Unlike the intergenic sequences previously described [25], large vertebrate genes do not have predominantly disordered CSDs. V-shaped CSDs (Fig. 1; Supplementary figure 1) are not the most common type except the CG skew diagrams of *D. rerio* and *R. norvegicus* (Table 1). Together with other evidence that cumulative CG skew diagrams are not as canonical as cumulative AT skew diagrams [25], it can be concluded that, in only a small proportion of genes, replication-associated strand-asymmetric substitutions are the sole or dominant sequence change. A large number of genes' CSDs are clear sloping lines, with a few being/-shaped, but most being\-shaped (Fig. 1; Supplementary figure 2).

A sloping CSD means that the strand asymmetry is constant along the gene. A \-shaped CSD means that there is a preference for T over A and G over C in the coding strand, which is equivalent to the transcription-associated strand compositional asymmetries previously reported [20–22, 26, 27]. Since the replicon sizes vary more than tenfold within each animal genome from less than 50 kb to more than 1 Mb [25, 33], a gene larger than 50 kb we selected may happen to reside within a single leading or lagging strand and so replication-associated strand-asymmetric substitutions should also be constant within the gene. Then, such a gene will have a \-shaped or a /-shaped CSD. We suggest that this may account for the /-shaped CSDs found and an equal number of the \-shaped CSDs. As shown in Table 1, even if the expected number of replication-associated \-shaped CSDs are subtracted from the total \-shaped CSDs, the transcription-associated \-shaped CSD is still the most common type.

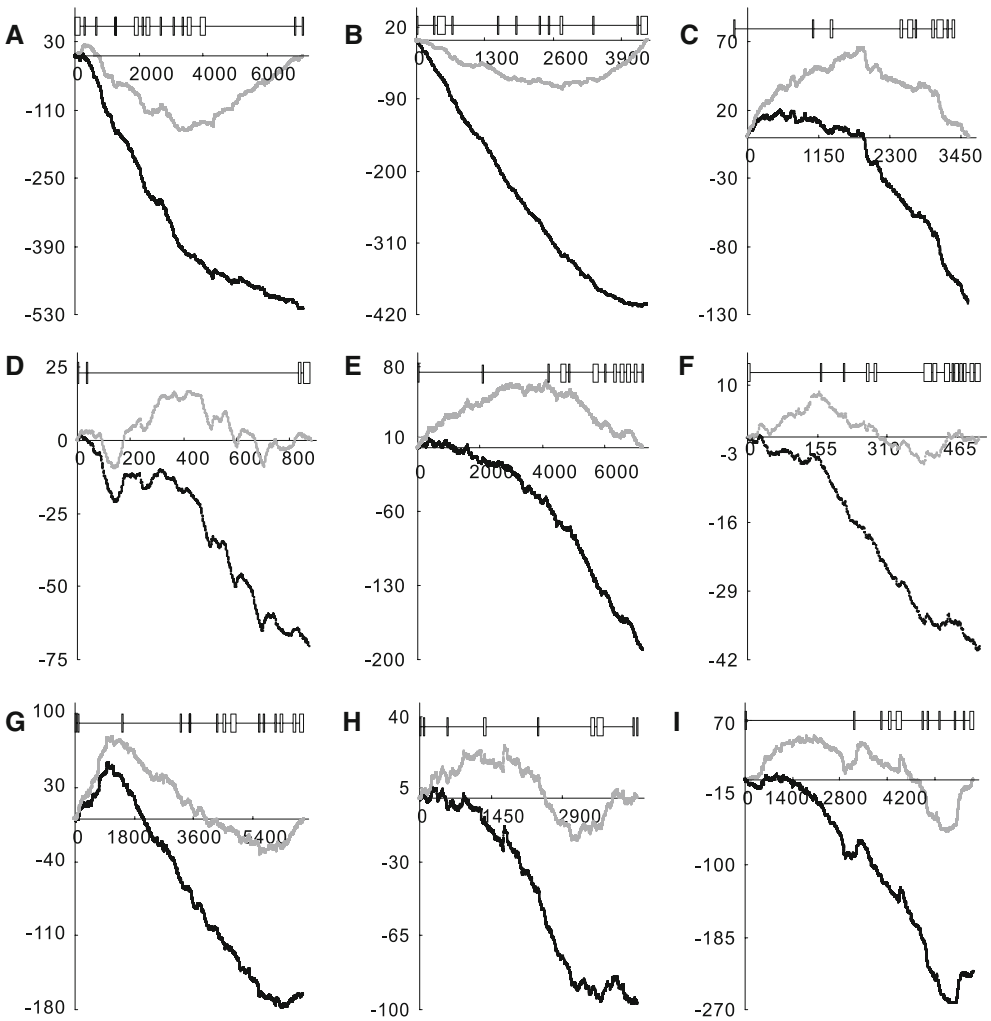

Fig. 2 Comparisons between cumulative skew diagrams (CSD) and cumulative skew difference diagrams (CSDD) of some examples of vertebrate large genes. The *X* axis represents the sequence length in 100 bp units, and the *Y* axis shows the cumulative skew values and cumulative skew difference values. The CSD are *black curves* and the CSDD are *grey curves*. Exons are shown by *boxes* or *vertical bars* (in the cases where exon is very small compared with the whole gene), while introns are shown by *horizontal lines*. The sizes of boxes and lines are proportional to the length of exon sizes and intron sizes. **a** AT diagrams of *Homo sapiens* gene *BCAS3*; **b** AT diagrams of *Homo sapiens* gene *SND1*; **c** AT diagrams of *Homo sapiens* gene *MEGF11*; **d** AT diagrams of *Mus musculus* gene *Zfp119*; **e** CG diagrams of *Gallus gallus* gene *LOC417729*; **f** AT diagrams of *Canis familiaris* gene *LOC490941*; **g** CG diagrams of *Canis familiaris* gene *LOC485651*; **h** CG diagrams of *Pan troglodytes* gene *LOC472962*; **i** CG diagrams of *Pan troglodytes* gene *LOC470764*

Table 2 Survey of the abundance of introns in the large genes we studied

Species	Number of genes studied	Intron length/exon length		Gene length/intron number	
		Mean ± SEM	Median	Mean ± SEM	Median
Danio rerio	1227	46.8 ± 1.2	35.3	9710 ± 306	6631
Gallus gallus	1205	44.7 ± 1.6	28.3	8982 ± 350	5340
Canis familiaris	1422	57.4 ± 1.7	37.7	11662 ± 468	6637
Mus musculus	1543	65.3 ± 2.2	40.5	18566 ± 684	10433
Rattus norvegicus	1598	67.7 ± 2.5	41.9	16271 ± 684	8305
Pan troglodytes	1982	92.0 ± 2.4	60.4	19610 ± 644	10864
Homo sapiens	1684	66.3 ± 2.4	40.6	20109 ± 855	10712

Table 3 Housekeeping genes have higher proportions of \-shaped cumulative skew diagrams (CSDs) than somatic-cell-expressed genes

Species	AT skew				CG skew			
	Number of \-shaped CSDs	Number of other CSDs	χ^2	P	Number of \-shaped CSDs	Number of other CSDs	χ^2	P
Homo sapiens								
Housekeeping genes	138	45	5.95	0.015	140	43	21.3	4.0×10^{-6}
Somatic-cell expressed genes	93	56			77	72		
Mus musculus								
Housekeeping genes	48	17	3.30	0.069	41	24	13.0	3.1×10^{-4}
Somatic-cell expressed genes	153	99			94	158		

Table 4 Housekeeping genes have stronger transcription-associated strand compositional asymmetry than somatic-cell expressed genes

	AT skew				CG skew			
	n	Mean ± SEM	Median	P^a	n	Mean ± SEM	Median	P^a
Homo sapiens								
Housekeeping genes with \-shaped CSDs	138	–0.071 ± 0.002	–0.069	3.5×10^{-3}	140	–0.036 ± 0.001	–0.034	6.1×10^{-4}
Somatic-cell expressed genes with \-shaped CSDs	93	–0.059 ± 0.003	–0.056		77	–0.029 ± 0.001	–0.028	
Mus musculus								
Housekeeping genes with \-shaped CSDs	48	–0.075 ± 0.004	–0.068	1.8×10^{-3}	41	–0.042 ± 0.003	–0.045	7.2×10^{-6}
Somatic-cell expressed genes with \-shaped CSDs	153	–0.061 ± 0.002	–0.056		94	–0.028 ± 0.002	–0.023	

[a] The data are normally distributed, so we used t test to calculated the P values

Coexistence of replication-associated and transcription-associated strand compositional asymmetries

To show whether replication-associated strand compositional asymmetries coexist with transcription-associated strand compositional asymmetries in the majority of genes, we used the CSDD to filter out the potential effects of transcription. V- (or multi-connected V-) shaped CSDDs were observed for most of the genes having \-shaped CSDs (Table 1). Figure 2 illustrates a comparison between CSDs and CSDDs for six vertebrate genes. The diagrams for more genes are deposited in Supplementary figure 3. These diagrams clearly show that transcription-associated strand compositional asymmetries are generally much stronger than, and thus can conceal the replication-associated strand compositional asymmetries. This may be the reason why traditional chromosomal-wide approaches failed to reveal replication-associated strand compositional asymmetries in eukaryotes [3, 4, 36, 37].

Transcription-associated strand asymmetries: mutation versus selection

Transcription-associated strand compositional asymmetries have been observed in a high percentage of genes (Table 1 and [21]), which are intuitively much more than that

required to be expressed in germline cells. Because only mutations occurring in the germ line can be transmitted to next generation and accumulate in evolution, it seems to support that selective pressures are the main cause of transcription-associated strand compositional asymmetries. But the gene length/intron number ratios are very large in all the species we studied (Table 2), splicing-associated selective pressures [27] unlikely contribute much. Meanwhile the introns are generally much longer than the exons in the large genes we studied (Table 2), which is also marked out in each diagram. The exons unlikely contribute much to the global shape of the diagrams (Figs. 1, 2 and Supplementary figures 1, 2, and 3). So selective pressures operating on exons (like codon bias and the existence of exonic splicing enhancers [28–30]) unlikely contribute much to the global strand compositional asymmetries of vertebrate large genes.

One the other hand, there is evidence that many (or even most) tissue-specific genes are ectopically transcribed in various cell types [38–40]. Niu proposed that this widely low-level ectopical transcriptions of tissue-specific genes may be to silence the genes by small RNA-mediated transcriptional repression [41]. The tissue-specific genes may be expressed at low-level in germline cells, so transcription-associated strand asymmetric mutations may account for the transcription-associated strand compositional asymmetries observed in most genes [21 and above results in present paper]. Nonetheless, ectopical germline transcription levels of somatic tissue-specific genes would be much lower than the transcription levels of housekeeping genes in germline cells. If mutational pressures are the main cause of transcription-associated strand compositional asymmetries, we would expect a clear difference in the asymmetries between somatic-cell-expressed genes and housekeeping genes. Actually, we found a higher percentage of housekeeping genes having \-shaped CSDs (Table 3). Further, the strand compositional asymmetries of the housekeeping genes having \-shaped CSDs are significantly stronger than those of somatic-cell-expressed genes with \-shaped CSDs (Table 4).

In summary, we found that the replication-associated strand compositional asymmetries are much weaker than transcription-associated ones in most genes, and provided evidence that the main cause of transcription-associated strand compositional asymmetry is transcription-associated strand asymmetric mutations (e.g. transcription-coupled repair [20] and transcription-coupled mutagenesis [26]), rather than transcription-associated strand asymmetric selection pressures.

Acknowledgments We thank Lei Zhu and Shu-Wei Li for technical helps. This work was supported by the National Natural Science Foundation of China (30270695) and Beijing Normal University.

References

1. Lobry JR (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. Mol Biol Evol 13(5):660–665
2. Francino MP, Ochman H (1997) Strand asymmetries in DNA evolution. Trends Genet 13(6):240–245
3. Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 26(10):2286–2290
4. Mrazek J, Karlin S (1998) Strand compositional asymmetry in bacterial and large viral genomes. Proc Natl Acad Sci USA 95(7):3720–3725
5. Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol 15(8):957–966
6. Frank AC, Lobry JR (1999) Asymmetric substitution patterns: A review of possible underlying mutational or selective mechanisms. Gene 238(1):65–77
7. Grigoriev A (1999) Strand-specific compositional asymmetries in double-stranded DNA viruses. Virus Res 60(1):1–19
8. Rocha EPC, Danchin A, Viari A (1999) Universal replication biases in bacteria Mol Microbiol 32(1):11–16
9. Lobry JR, Sueoka N (2002) Asymmetric directional mutation pressures in bacteria. Genome Biol 3(10): research0058.0051–0058.0014
10. Tillier ERM, Collins RA (2000) The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J Mol Evol 50(3):249–257
11. Kowalczuk M, Mackiewicz P, Mackiewicz D, Nowicka A, Dudkiewicz M, Dudek MR, Cebrat S (2001) DNA asymmetry and the replicational mutational pressure. J Appl Genet 42(4):553–577
12. Beletskii A, Bhagwat AS (1998) Correlation between transcription and C to T mutations in the non-transcribed DNA strand. Biol Chem 379(4–5):549–551
13. Francino MP, Ochman H (2001) Deamination as the basis of strand-asymmetric evolution in transcribed *Escherichia coli* sequences. Mol Biol Evol 18(6):1147–1150
14. Zeigler DR, Dean DH (1990) Orientation of genes in the *Bacillus subtilis* chromosome. Genetics 125(4):703–708
15. McLean MJ, Wolfe KH, Devine KM (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J Mol Evol 47(6):691–696
16. Rocha EP, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat Genet 34(4):377–378
17. Baran RH, Ko H (2006) An Ising model of transcription polarity in bacterial chromosomes. Physica A 362(2):403–422
18. Nikolaou C, Almirantis Y (2005) 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 33(21):6816–6822
19. Baran RH, Ko H, Jernigan RW (2003) Methods for comparing sources of strand compositional asymmetry in microbial chromosomes. DNA Res 10(3):85–95
20. Green P, Ewing B, Miller W, Thomas PJ, Green ED (2003) Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet 33(4):514–517
21. Majewski J (2003) Dependence of mutational asymmetry on gene-expression levels in the human genome. Am J Hum Genet 73(3):688–692
22. Niu DK, Lin K, Zhang DY (2003) Strand compositional asymmetries of nuclear DNA in eukaryotes. J Mol Evol 57(3):325–334
23. Touchon M, Nicolay S, Audit B, Brodie of Brodie E-B, d'Aubenton-Carafa Y, Arneodo A, Thermes C (2005) Replication-associated strand asymmetries in mammalian genomes: toward

detection of replication origins. Proc Natl Acad Sci USA 102(28):9836–9841
24. Brodie of Brodie E-B, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Thermes C, Arneodo A (2005) From DNA sequence analysis to modeling replication in the human genome. Phys Rev Lett 9424(24):248103
25. Hou WR, Wang HF, Niu DK (2006) Replication-associated strand asymmetries in vertebrate genomes and implications for replicon size, DNA replication origin, and termination. Biochem Biophys Res Commun 344(4):1258–1262
26. Touchon M, Nicolay S, Arneodo A, d'Aubenton-Carafa Y, Thermes C (2003) Transcription-coupled TA and GC strand asymmetries in the human genome. FEBS Lett 555(3):579–582
27. Touchon M, Arneodo A, d'Aubenton-Carafa Y, Thermes C (2004) Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes. Nucleic Acids Res 32(17):4969–4978
28. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet 7(2):98–108
29. Parmley JL, Chamary JV, Hurst LD (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol 23(2):301–309
30. Pozzoli U, Riva L, Menozzi G, Cagliani R, Comi GP, Bresolin N, Giorda R, Sironi M (2004) Over-representation of exonic splicing enhancers in human intronless genes suggests multiple functions in mRNA processing. Biochem Biophys Res Commun 322(2): 470–476
31. Lobry JR (1996) A simple vectorial representation of DNA sequences for the detection of replication origins in bacteria. Biochimie 78(5):323–326
32. Zhang R, Zhang C-T (2005) Identification of replication origins in archaeal genomes based on the Z-curve method. Archaea 1(5):335–346
33. Lewin B (2004) Genes, VIII edn. Pearson Prentice Hall, Upper Saddle River
34. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G et al (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 101(16):6062–6067
35. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A et al (2002) Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA 99(7):4465–4470
36. Karlin S, Campbell AM, Mrazek J (1998) Comparative DNA analysis across diverse genomes. Annu Rev Genet 32:185–225
37. Gierlik A, Kowalczuk M, Mackiewicz P, Dudek MR, Cebrat S (2000) Is there replication-associated mutational pressure in the *Saccharomyces cerevisiae* genome? J Theor Biol 202(4):305–314
38. Chelly J, Concordet JP, Kaplan JC, Kahn A (1989) Illegitimate transcription: transcription of any gene in any cell type. Proc Natl Acad Sci USA 86(8): 2617–2621
39. Kimoto Y (1998) A single human cell expresses all messenger ribonucleic acids: the arrow of time in a cell. Mol Gen Genet 258(3):233–239
40. Sarkar G, Sommer SS (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. Science 244(4902):331–334
41. Niu D-K (2005) Low-level illegitimate transcription of genes may be to silence the genes. Biochem Biophys Res Commun 337(2):413–414