View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Springer - Publisher Connector

BMC Genomics

BioMed Central

Research article

Open Access

Evolutionary implications of inversions that have caused intra-strand parity in DNA

Kohji Okamura[1,2], John Wei[1] and Stephen W Scherer*[1,2]

Address: [1]The Centre for Applied Genomics, Program in Genetics and Genome Biology, The Hospital for Sick Children, MaRS Centre, Toronto, Ontario, Canada and [2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada

Email: Kohji Okamura - kohji@genet.sickkids.on.ca; John Wei - wei@genet.sickkids.on.ca; Stephen W Scherer* - steve@genet.sickkids.on.ca

* Corresponding author

Published: 11 June 2007

BMC Genomics 2007, **8**:160 doi:10.1186/1471-2164-8-160

Received: 25 January 2007

Accepted: 11 June 2007

This article is available from: http://www.biomedcentral.com/1471-2164/8/160

© 2007 Okamura et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Chargaff's rule of DNA base composition, stating that DNA comprises equal amounts of adenine and thymine (%A = %T) and of guanine and cytosine (%C = %G), is well known because it was fundamental to the conception of the Watson-Crick model of DNA structure. His second parity rule stating that the base proportions of double-stranded DNA are also reflected in single-stranded DNA (%A = %T, %C = %G) is more obscure, likely because its biological basis and significance are still unresolved. Within each strand, the symmetry of single nucleotide composition extends even further, being demonstrated in the balance of di-, tri-, and multi-nucleotides with their respective complementary oligonucleotides.

Results: Here, we propose that inversions are sufficient to account for the symmetry within each single-stranded DNA. Human mitochondrial DNA does not demonstrate such intra-strand parity, and we consider how its different functional drivers may relate to our theory. This concept is supported by the recent observation that inversions occur frequently.

Conclusion: Along with chromosomal duplications, inversions must have been shaping the architecture of genomes since the origin of life.

Background

The most famous of Chargaff's rules is that in DNA, the proportion of A equals that of T, and C that of G [1]. This nucleotide balance is governed by complementary base-pairing rules fundamental to the structure of the double helix [2]. Astonishingly, the nucleotides retain almost the same equality balance in either of the two single strands of DNA [3] and this phenomenon is sometimes named Chargaff's second parity rule [4-10]. Table 1 provides an illustration, with analysis of large contiguous segments from each human chromosome.

When there is no bias in mutation and selection between complementary strands, base substitution may explain the parity phenomenon [11,12]. In fact, strand bias has been demonstrated with mutational skews between the two strands, which causes deviation from parity [13,15]. Bacterial origins of replication were successfully identified by the distribution of such skews [16,17]. The strand bias of mutations, which can be associated with direction of transcription, is also found in mammalian genomes [18,19]. In spite of these anomalies, any violation of the second parity phenomena is generally small in magnitude [8,20].

Table 1: Mononucleotide content in contiguous single-stranded DNA scaffolds from each human chromosome *

Chromosome	Accession number	Length	%A	%T	%C	%G
1	NT_032977	73,835,825	29.72	29.69	20.33	20.27
2	NT_005403	84,213,157	30.60	30.68	19.34	19.38
3	NT_005612	100,530,253	30.51	30.53	19.46	19.49
4	NT_016354	92,123,751	31.34	31.33	18.64	18.69
5	NT_006576	46,378,398	30.45	30.31	19.62	19.62
6	NT_025741	61,645,385	30.84	30.86	19.16	19.14
7	NT_007933	64,426,257	30.43	30.39	19.62	19.56
8	NT_008046	57,155,273	30.21	30.04	19.89	19.86
9	NT_008470	40,394,265	28.72	28.72	21.27	21.28
10	NT_030059	44,617,998	29.12	29.30	20.80	20.77
11	NT_009237	49,571,094	29.57	29.70	20.36	20.37
12	NT_029419	38,648,979	30.06	30.01	19.96	19.97
13	NT_024524	67,740,325	30.97	30.93	19.06	19.04
14	NT_026437	88,290,585	29.44	29.67	20.42	20.47
15	NT_010194	53,619,965	29.06	28.82	21.11	21.01
16	NT_010498	42,003,582	28.32	28.31	21.66	21.70
17	NT_010783	24,793,602	28.22	28.25	21.76	21.76
18	NT_010966	33,548,238	30.34	30.23	19.73	19.71
19	NT_011109	31,383,029	26.25	26.32	23.68	23.76
20	NT_011362	26,144,333	27.26	27.56	22.57	22.61
21	NT_011512	28,617,429	30.57	30.31	19.60	19.52
22	NT_011520	23,276,302	26.33	26.29	23.72	23.67
X	NT_011651	36,813,576	31.07	31.36	18.74	18.82
Y	NT_011875	10,002,238	30.43	30.52	19.35	19.70
mtDNA	NC_001807	16,571	30.86	24.66	31.33	13.16

* The longest contig was chosen from each human chromosome.

Although different explanations for this parity phenomenon have been put forth, such as intra-strand base pairing [6], a simpler explanation for the rule may be DNA duplication and inversion [4,8,10]. If double-stranded DNA of any composition undergoes duplication followed by an inversion of the duplicated region, then each strand of the resulting DNA molecule would precisely satisfy Chargaff's second parity rule, so that %A = %T and %C = %G (Fig. 1A).

Not only single nucleotides but also oligonucleotides up to 30 nucleotides (nt) in length can demonstrate the parity phenomenon within strands [5,7,8]. In other words, the frequency of a particular oligonucleotide is approximately equal to that of its reverse complementary sequence in the same strand. Since DNA strands are complementary, the frequency of a particular oligonucleotide in one strand approximates that in the opposite strand. Hence, this double-stranded DNA characteristic can also be called "symmetry of complementary DNA strands" [5,8]. Chargaff's second parity rule ordinarily considers only mononucleotides, which have been extensively studied. However, since a single nucleotide could be deemed a one-nt oligonucleotide, it is plausible that addressing the symmetry of oligonucleotides (high-order strand symmetry) is a more general way of assessing biological meaning. Hereafter, we designate this comprehensive symmetry as "intra-strand parity" and attempt to explain it based on the mechanism of chromosomal inversion. Single nucleotide mutations may be considered to explain mononucleotide parity within strands [11,12] but have not been effective to explain the extended parity of oligonucleotides [8].

Results

We propose that inversion events (with or without underlying duplications) might be a sufficient mechanism to explain the phenomenon. To test this, we consider a double-stranded DNA molecule without intra-strand parity but which is long enough to undergo various (stochastic) inversions (Fig. 1B). A_n and T_n are defined as the frequency of any particular oligonucleotide sequence and its reverse complementary sequence, respectively, in the same strand after n inversions ($n > 0$). A_0 ($0 < A_0 < 1$) is the initial frequency of any particular oligonucleotide sequence (which can also be a mononucleotide) in the upper strand. T_0 ($0 < T_0 < 1$) is the initial frequency of its reverse complementary sequence in the same strand. If we define r_n ($0 < r_n << 1$) as the relative length of the nth inversion (Fig. 1B), we obtain these two equations.

$$A_n = A_{n-1} - r_n(A_{n-1} - T_{n-1}) \quad (1)$$

$$T_n = T_{n-1} - r_n(T_{n-1} - A_{n-1}) \quad (2)$$

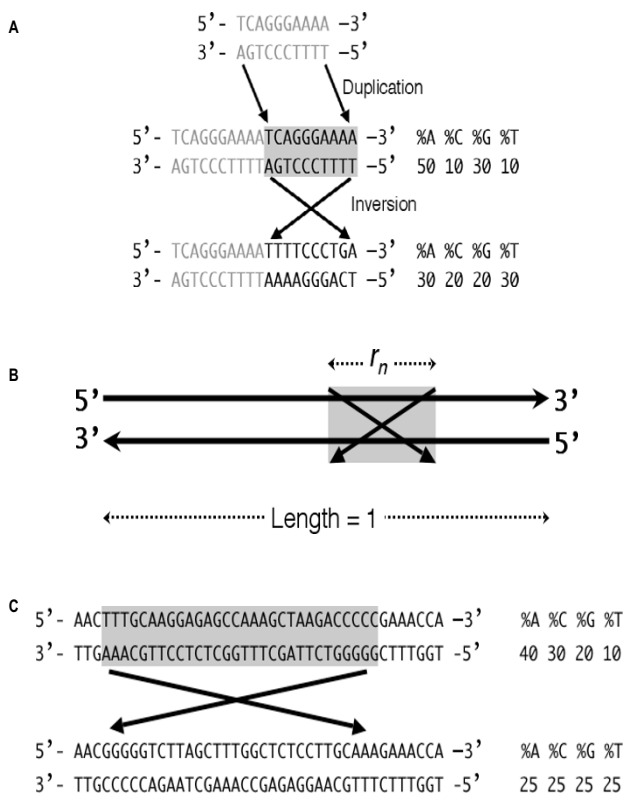

Figure 1

Inversions as an explanation for intra-strand parity. **A**, Duplication followed by inversion. If a double-stranded DNA, shown in gray, undergoes duplication and inversion, then the resulting molecule precisely demonstrates the strand parity (both within and between strands). **B**, A mathematical explanation of intra-strand parity. The *n*th inversion is illustrated by a box with crossed bars and r_n is the relative length of the inversion within a total fragment of length = 1. Ultimately both A_n and T_n converge to the average of their initial frequencies. See Methods for details. Although a linear double-stranded DNA is shown, this could also be circular. **C**, A small number of inversions can cause DNA to follow the intra-strand parity. A 40-bp double-stranded DNA fragment in the human mtDNA (position 1875–1914 in accession number NC_001807) is shown, along with the outcome of a single artificial inversion, which has homogenized the contents of the two strands.

Equations (1) and (2) mean that an inversion changes A_n and T_n toward T_n and A_n, respectively. When the whole sequence is long enough, r_n is close to 0. Nevertheless, whatever the size of the inverted region examined, any oligonucleotide sequence will eventually be homogenized between two strands. In other words, A_n and T_n ultimately converge to be equal to each other, regardless of r_n, as long as r_n is stochastic (see mathematical derivation in Methods).

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} T_n = \frac{A_0 + T_0}{2} \quad (3)$$

Equation (3) is a mathematical explanation of intra-strand parity based on our hypothesis that inversions are sufficient to cause any DNA segment conform to parity. In this way, the vast majority of naturally occurring DNA molecules (chromosomes) will evolve to intra-strand parity via many inversions. Those few that deviate, such as mitochondrial DNA (mtDNA) [8,9,17], will have special properties (see below). We presume that any DNA can be made to evolve to intra-strand parity through a process of inversions, and that deviations from parity have been rare in evolution. Inversions must have been occurring as genomes of ancestral organisms were growing in complexity with the acquisition or creation of new genes.

The insertion of repetitive sequences was proposed to be a possible source underlying parity [8,10]. However, removing apparent repeats from the human and other genomes prior to analysis (see Methods) did not alter the symmetry characteristics of the remaining sequences. (An example of a 28.6-Mb contig from human chromosome 21 is shown in Table 2). Therefore, it is unlikely that insertion of such sequences accounts for the intra-strand parity, either in humans or organisms that have fewer repetitive sequences in their genomes.

We employ radar charts to allow simple visual perception of the high-order symmetry and asymmetry of exemplary DNAs (Fig. 2). Mitochondria are thought to have been derived from bacteria [21]. Mammalian mtDNA (Fig 2C) is an exception that does not demonstrate intra-strand parity [8,9,17] whereas mtDNAs from plants and lower eukaryotes do. Mammalian mtDNA may have gradually deviated from its ancestral form [9]. The small circular size, its unique replication mechanism [22], and extranuclear localization could introduce different selective pressures against tolerance of inversions and thus deviation from the more general observation of intra-strand parity.

The mammalian mtDNA offers a natural source of sequence sufficiently deviating from parity to allow us to further test our mathematical explanation. We produced *in silico* semi-random inversions in human mtDNA. As few as eight 1-kb regularly-distributed inversions (see Methods) would be sufficient to homogenize the two strands of the 16.6-kb mtDNA and create intra-strand parity (Fig. 2D). We also depict a hypothetical inversion in the mtDNA to show the potential for rapid homogenization (Fig. 1C).

Table 2: Dinucleotide frequencies in a human genomic contig without repetitive sequences *

Dinucleotide	Frequency	Difference	Frequency	Dinucleotide
AA	0.10956	0.00084	0.10872	TT
AC	0.04992	0.00047	0.04945	GT
AG	0.06718	0.00016	0.06702	CT
AT	0.08639	0.00000	0.08639	AT
CA	0.07012	0.00072	0.06940	TG
CC	0.04309	0.00027	0.04282	GG
CG	0.00781	0.00000	0.00781	CG
CT	0.06702	0.00016	0.06718	AG
GA	0.05869	0.00008	0.05876	TC
GC	0.03630	0.00000	0.03630	GC
GG	0.04282	0.00027	0.04309	CC
GT	0.04945	0.00047	0.04992	AC
TA	0.07474	0.00000	0.07474	TA
TC	0.05876	0.00008	0.05869	GA
TG	0.06940	0.00072	0.07012	CA
TT	0.10872	0.00084	0.10956	AA
Total	1.00000		1.00000	Total

* See text and Methods.

Although the lack of intra-strand parity in mammalian mtDNA could be ascribed to its small length, other loci of comparable length (*e.g.* the *TP53* gene, Fig. 2B) do adhere to parity. Unlike other mtDNAs, those of mammals have no intergenic segments and have only one regulatory region per strand. Moreover, unlike among nuclear genomes, the order and direction of genes – as well as biased gene density between the two strands – are strictly conserved among mammalian species [23]. Therefore, it seems that the configuration is already fixed, and that inversions are not tolerated in mammalian mtDNA.

Discussion

The ubiquity of inversions suggests that they had some advantage in natural selection. Duplications are thought to play an important role in creating genetic variety [24], however, some duplications are deleterious for organisms, due to sudden increases of gene dosage. To avoid being negatively selected, one of the duplicated copies could undergo mutation such as deletion. Inversions or interchromosomal rearrangements could render the duplicated gene nonfunctional due to its release from interaction with its promoter or other regulatory elements. This may be one reason why many inverted and interchromosomal segmental duplications are found in the human genome [25,26]. An approximately symmetrical gene distribution between the two strands may have been brought about by these rearrangements [27].

In some cases, a rearranged genome might confer positive selection. Although we can find syntenic regions among vertebrates, chromosomal organizations can be quite different among species. This suggests an advantage for evolution or speciation. Recently, the importance of gene order and gene position in the three-dimensional nucleus has been suggested [28]. It is likely that genomes continually undergo rearrangement toward optimal positions for each gene and each gene cluster. Our group showed an unexpectedly large number of inversions (from 23 bp to 62 Mb in size) between human and chimpanzee genomes [29], species which diverged only six million years ago. Although most may be selectively neutral, some likely were selected for, and contributed to the speciation. Many more inversions may also have occurred and may have been negatively selected. Inversions can also give rise to new transcripts, some of which will be selected for and become new genes. We identified hybrid transcripts of the *AZGP1* and *GJE1* genes on human chromosome 7 (manuscript in preparation) and are intrigued that the orthologues of these genes in non-primate mammals reside in a head-to-head manner. It is likely that the common ancestor of primates underwent inversion of the *AZGP1* gene to produce the hybrid transcripts, creating an opportunity for primate diversity.

Conclusion

In summary, we propose that the relatively frequent occurrence and accumulation of inversions in genomes may be a major contributor to the phenomenon of intra-strand parity. Whereas single base substitutions might explain Chargaff's second parity rule at the level of mononucleotides, they can explain neither the high-order intra-strand parity nor the exceptional deviation of mammalian mtDNAs. In contrast, inversion events are not limited by size and can involve millions of bases of sequence. Other mechanisms may have contributed to some extent; nevertheless, they are not necessary to account for intra-strand parity if inversions are considered.

Inversions are one process contributing to genome evolution that allow for rearrangement toward optimal position, order, and orientation of genes and regulatory elements, and for escape from deleterious effects caused, for example, by some duplications. Although we acknowledge the possibility of preferential sites, inversions occur randomly as shown in our mathematical explanation. Many of these are expected to be deleterious and would presumably be selected against, but others should be neutral or positively selected and could therefore become fixed in the genome [30]. Quantitative estimation of inversion using genomic sequences of extant organisms is unfortunately meaningless, as it cannot account for those events lost to natural selection. Further, inversions must have contributed to the basic character of DNA sequences since the origin of life. There are now substantial data supporting the frequency of inversions within genomes of a variety of organisms, including plants, insects and pri-

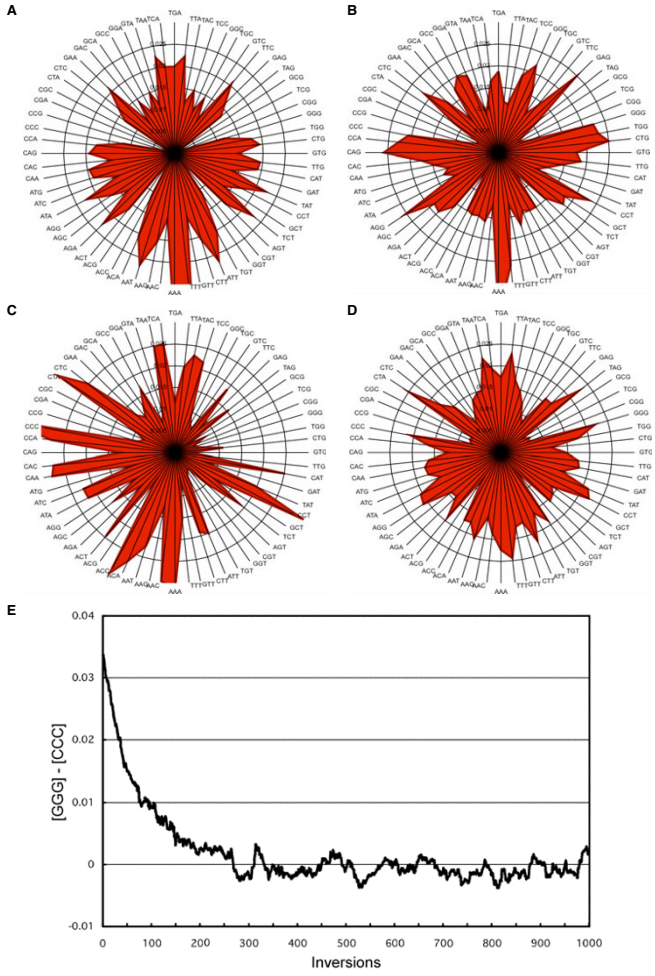

Figure 2

Intra-strand parity visually represented by radar charts. Frequencies of trinucleotides in various DNA sequences are shown here. Each trinucleotide is sorted alphabetically from bottom to top (left side). The corresponding complementary trinucleotides are arranged across to the right. **A**, Radar chart representing a fully sequenced contig (NT_010966, 33,548,238 bp) of human chromosome 18. This contig is continuous and does not include any annotated gaps or ambiguous nucleotides. The symmetrical chart shows the equal frequencies of specific oligonucleotides and their reverse complementary oligonucleotides. The high frequencies of poly-A and poly-T, which might be, in part, traces of retrotranspositions of poly-A^+ mRNA, and the deficiencies of trinucleotides that contain the CpG dinucleotide make the stalk and four grooves, respectively, of the "maple leaf" shape. (The shapes vary slightly based on the genome sequence analyzed, but the general symmetry is maintained). **B**, The genomic sequence of the *p53* (*TP53*) locus (U94788, 20,303 bp). The symmetry is roughly retained in sequences as short as 20 kb in length. The protein-coding sequences occupy 5.8% of this locus. This chart also suggests that transcriptional asymmetry is small in magnitude. **C**, Human mtDNA. The asymmetry illustrates that this DNA does not show intra-strand parity. **D**, Human mtDNA after inversion *in silico*. It becomes symmetrical, demonstrating that inversions can change a sequence to create the parity. In this case, each r_n approximates to 1/16.6. This also demonstrates that only $1/(2r_{ave})$ inversions (eight inversions in this case) are enough to make a sequence conform to parity. **E**, The difference of frequencies of GGG and CCC ([GGG] - [CCC]) in human mtDNA approaches 0 by *in silico* random inversions. In this analysis, for simplicity, the size of each inversion was fixed to 100 bp. In human mtDNA, GGG and CCC have the largest difference of frequencies among all trinucleoties (see Fig. 2C).

mates [29-33], and these observable events are but the tip of the iceberg. Chromosomal rearrangements such as inversions reduce the rate of meiotic recombination between homologous chromosomes, with subsequent reproductive isolation [34]. Moreover, in these regions, mutations tend to be positively selected to give rise to speciation [35]. Ohno's seminal work [24] and that of others have emphasized the importance of duplications in evolution. Our suppositions further these ideas, in particular suggesting how inversions and duplications can complement each other to yield the properties of extant genomes.

Methods

Calculation of frequencies of oligonucleotides

The genomic sequences (human contigs, the *TP53* gene, and the mtDNA sequence) were downloaded from NCBI (Build 36). Calculation of frequencies of oligonucleotides (including mononucleotides) was performed using Perl scripts, which are available upon request. The "plus" strand, which is stored in the database, was analyzed. We generated sequence free of repetitive elements using RepeatMasker with which 46.4% of the 28,617,429 nucleotides were masked. The coordinates of the eight 1-kb regularly-scattered *in silico* inversions were 1001–2000, 3001–4000, 5001–6000, 7001–8000, 9001–10000, 11001–12000, 13001–14000, and 15001–16000 in NC_001807.

Mathematical derivation

For the frequency of a particular oligonucleotide A_n ($n > 0$), via the nth inversion, $(1 - r_n)\ A_{n-1}$ remains; $r_n A_{n-1}$ decreases; $r_n T_{n-1}$ increases if we suppose the distribution of contents is even in the whole sequence. In this way, the two recurrence formulas (1) and (2) are derived (see text). The following equations are obtained by adding equations (1) and (2).

$$A_n + T_n = A_{n-1} + T_{n-1} \quad (4)$$

$$A_n + T_n = A_0 + T_0 \quad (5)$$

These mean that inversions do not change the sum of the two frequencies. Using (5), other forms of (1) and (2) are derived.

$$A_n = (1 - 2r_n)A_{n-1} + r_n(A_0 + T_0) \quad (6)$$

$$T_n = (1 - 2r_n)T_{n-1} + r_n(A_0 + T_0) \quad (7)$$

When we subtract $(A_0 + B_0)/2$ from (6) and define B_n, (9) is derived.

$$B_n = A_n - \frac{A_0 + B_0}{2} \quad (8)$$

$$\begin{aligned} B_n &= (1-2r_n)B_{n-1} \\ &= (1-2r_1)(1-2r_2)(1-2r_3)\ldots(1-2r_{n-1})B_0 \\ &= B_0\prod_{k=1}^{n-1}(1-2r_k) \\ &= \frac{A_0 - T_0}{2}\prod_{k=1}^{n-1}(1-2r_k) \end{aligned} \quad (9)$$

Using $-1 << 1 - 2r_k < 1$ $(0 < r_k << 1)$,

$$\lim_{n\to\infty} B_n = \frac{A_0 - T_0}{2}\lim_{n\to\infty}\prod_{k=1}^{n-1}(1-2r_k) = 0\,.$$

Therefore, $\lim_{n\to\infty} A_n = \lim_{n\to\infty} B_n + \frac{A_0 + T_0}{2} = \frac{A_0 + T_0}{2}$.

Similarly, $\lim_{n\to\infty} T_n = \frac{A_0 + T_0}{2}$.

Authors' contributions

KO conceived the study, performed the computational analyses, mathematical derivation, and drafted the manuscript. JW participated in the coordination of the study and performed the computational analyses. SWS participated in the design and coordination of the study and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank J. Buchanan, O. Akiyama, S. Horike, C. R. Marshall, A. Navarro, P. Pevzner, R. F. Wintle and J. Zhang for discussions and critical reading of the manuscript. We acknowledge the Centre for Computational Biology and The Centre for Applied Genomics for computational assistance. The work is supported by Genome Canada/Ontario Genomics Institute, the McLaughlin Centre for Molecular Medicine, and The Hospital for Sick Children Foundation. S.W.S. is an Investigator of the Canadian Institutes for Health Research and International Scholar of the Howard Hughes Medical Institute.

References

1. Chargaff E: **Structure and function of nucleic acids as cell constituents.** *Fed Proc* 1951, **10:**654-659.
2. Watson JD, Crick FH: **Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid.** *Nature* 1953, **171:**737-738.
3. Rudner R, Karkas JD, Chargaff E: **Separation of *B. subtilis* DNA into complementary strands. 3. Direct analysis.** *Proc Natl Acad Sci USA* 1968, **60:**921-922.
4. Fickett JW, Torney DC, Wolf DR: **Base compositional structure of genomes.** *Genomics* 1992, **13:**1056-1064.
5. Prabhu VV: **Symmetry observations in long nucleotide sequences.** *Nucleic Acids Res* 1993, **21:**2797-2800.
6. Forsdyke DR, Mortimer JR: **Chargaff's legacy.** *Gene* 2000, **261:**127-137.
7. Qi D, Cuticchia AJ: **Compositional symmetries in complete genomes.** *Bioinformatics* 2001, **17:**557-559.
8. Baisnée PF, Hampson S, Baldi P: **Why are complementary DNA strands symmetric?** *Bioinformatics* 2002, **18:**1021-1033.

9. Mitchell D, Bridge R: **A test of Chargaff's second rule.** *Biochem Biophys Res Commun* 2006, **340:**90-94.
10. Albrecht-Buehler G: **Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions.** *Proc Natl Acad Sci USA* 2006, **103:**17828-17833.
11. Sueoka N: **Intrastrand parity rules of DNA base composition and usage biases of synonymous codons.** *J Mol Evol* 1995, **40:**318-325.
12. Lobry JR: **Properties of a general model of DNA evolution under no-strand-bias conditions.** *J Mol Evol* 1995, **40:**326-330.
13. McLean MJ, Wolfe KH, Devine KM: **Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes.** *J Mol Evol* 1998, **47:**691-696.
14. Bell SJ, Forsdyke DR: **Deviations from Chargaff's second parity rule correlate with direction of transcription.** *J Theor Biol* 1999, **197:**63-76.
15. Daubin V, Perriere G: **G+C3 structuring along the genome: a common feature in prokaryotes.** *Mol Biol Evol* 2003, **20:**471-483.
16. Nikolaou C, Almirantis Y: **A study on the correlation of nucleotide skews and the positioning of the origin of replication: different modes of replication in bacterial species.** *Nucleic Acids Res* 2005, **33:**6816-6822.
17. Nikolaou C, Almirantis Y: **Deviations from Chargaff's second parity rule in organellar DNA insights into the evolution of organellar genomes.** *Gene* 2006, **381:**34-41.
18. Green P, Ewing B, Miller W, Thomas PJ, NISC Comparative Sequencing Program, Green ED: **Transcription-associated mutational asymmetry in mammalian evolution.** *Nat Genet* 2003, **33:**514-517.
19. Louie E, Ott J, Majewski J: **Nucleotide frequency variation across human genes.** *Genome Res* 2003, **13:**2594-2601.
20. Prescott DM, Dizick SJ: **A unique pattern of intrastrand anomalies in base composition of the DNA in hypotrichs.** *Nucleic Acids Res* 2000, **28:**4679-4688.
21. Fileé J, Forterre P: **Viral proteins functioning in organelles: a cryptic origin?** *Trends Microbiol* 2005, **13:**510-513.
22. Clayton DA: **Replication of animal mitochondrial DNA.** *Cell* 1982, **28:**693-705.
23. Pääbo S, Thomas WK, Whitfield KM, Kumazawa Y, Wilson AC: **Rearrangements of mitochondrial transfer RNA genes in marsupials.** *J Mol Evol* 1991, **33:**426-430.
24. Ohno S: *Evolution by Gene and Genome Duplication* Springer, Berlin; 1970.
25. 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.
26. Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW: **Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence.** *Genome Biol* 2003, **4:**R25.
27. Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, Ainscough R, Almeida JP, Babbage A, Bagguley C, Bailey J, Barlow K, Bates KN, Beasley O, Bird CP, Blakey S, Bridgeman AM, Buck D, Burgess J, Burrill WD, O'Brien KP, *et al.*: **The DNA sequence of human chromosome 22.** *Nature* 1999, **402:**489-495.
28. Kosak ST, Groudine M: **Gene order and dynamic domains.** *Science* 2004, **306:**644-647.
29. Feuk L, MacDonald JR, Tang T, Carson AR, Li M, Rao G, Khaja R, Scherer SW: **Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies.** *PLoS Genet* 2005, **1:**e56.
30. Hoffmann AA, Sgrò CM, Weeks AR: **Chromosomal inversion polymorphisms and adaptation.** *Trends Ecol Evol* 2004, **19(9):**482-488.
31. Blanc G, Barakat A, Guyot R, Cooke R, Delseny M: **Extensive duplication and reshuffling in the *Arabidopsis* genome.** *Plant Cell* 2000, **12:**1093-1101.
32. Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V: **A polytene chromosome analysis of the *Anopheles gambiae* species complex.** *Science* 2002, **298:**1415-1418.
33. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE: **Fine-scale structural variation of the human genome.** *Nat Genet* 2005, **37:**727-732.
34. Rieseberg LH: **Chromosomal rearrangements and speciation.** *Trends Ecol Evol* 2001, **16(7):**351-358.
35. Navarro A, Barton NH: **Chromosomal speciation and molecular divergence – accelerated evolution in rearranged chromosomes.** *Science* 2003, **300:**321-324.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral