Biochemical and Biophysical Research Communications 409 (2011) 367–371

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Mini Review

Exploring the common molecular basis for the universal DNA mutation bias: Revival of Löwdin mutation model

Liang-Yu Fu [a,b], Guang-Zhong Wang [c], Bin-Guang Ma [b], Hong-Yu Zhang [a,b,*]

[a] National Key Laboratory of Crop Genetic Improvement, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China
[b] Center for Bioinformatics, Huazhong Agricultural University, Wuhan 430070, PR China
[c] State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, PR China

ARTICLE INFO

Article history:
Received 26 April 2011
Available online 10 May 2011

Keywords:
Mutation
Oxidative damage
UV-radiation damage
CpG hypermutation
Löwdin model

ABSTRACT

Recently, numerous genome analyses revealed the existence of a universal G:C → A:T mutation bias in bacteria, fungi, plants and animals. To explore the molecular basis for this mutation bias, we examined the three well-known DNA mutation models, i.e., oxidative damage model, UV-radiation damage model and CpG hypermutation model. It was revealed that these models cannot provide a sufficient explanation to the universal mutation bias. Therefore, we resorted to a DNA mutation model proposed by Löwdin 40 years ago, which was based on inter-base double proton transfers (DPT). Since DPT is a fundamental and spontaneous chemical process and occurs much more frequently within GC pairs than AT pairs, Löwdin model offers a common explanation for the observed universal mutation bias and thus has broad biological implications.

DNA mutation is one of the most fundamental forces to drive biological evolution and to induce various diseases, such as cancer and inherited diseases. Elucidating the molecular mechanisms underlying DNA mutation is thus a very challenging topic in genetics and evolutionary biology. Recently, numerous analyses revealed the existence of a universal DNA mutation bias, namely, G:C → A:T mutations are preferred to A:T → G:C counterparts, in bacteria, fungi, plants and animals [1–10]. Since this biased mutation can be observed under weak selective pressure, it seems to be an inherent property of DNA. Therefore, it is of great interest and significance to explain this mutation bias at molecular level. In particular, it is challenging to explore whether there exists a common basis for this universal mutation bias.

A straightforward explanation to this universal mutation bias arises from the direct damage of reactive oxygen species (ROS) on DNA. As known to all, G and C can be readily transformed to 8-oxoguanine and 5-hydroxyuracil through attack of hydroxyl radical and oxidative deamination, respectively, which then causes G:C → A:T transitions and G:C → T:A transversions [11,12]. This mechanism has been used by Denver et al. to explain the much higher average conditional mutation rates of G:C → A:T transitions and G:C → T:A transversions in *Caenorhabditis elegans* [6]. Since oxygen is indispensable to ROS damage, to explore whether this mechanism is applicable to the universal DNA mutation bias, we analyzed the dependence of the mutation bias on the oxygen abundance.

In a recent work, Hildebrand and co-workers analyzed the mutation patterns of 4-fold sites of synonymous codons for 149 phylogenetically diverse bacteria and found that the proportions of G:C → A:T mutations in G:C ↔ A:T mutations (Z values) are usually larger than the predicted Z values (Z_{pred}), that is, Z–Z_{pred} is positive, which means an excess of G:C → A:T mutations over A:T → G:C counterparts [9]. Through searching NCBI Entrez Genome Project database (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi), we determined the oxygen requirement characteristics for the 149 bacterial species, in which 120 bacteria are aerobic (including strictly and facultative aerobic and microaerophilic) and 16 are strictly anaerobic (Table S1). As shown in Table S2, there is no significant difference between the average values of Z–Z_{pred} for aerobic and anaerobic bacteria ($P = 0.0827$, Wilcoxon two-sample test). In both types of bacteria, the proportions of members with positive Z–Z_{pred} are also comparable ($P = 0.1082$, Chi-square test). Therefore, it seems that for bacteria there is no close correlation between the mutation bias and oxygen requirement, which implies that ROS damage is unlikely to be the molecular basis for the universal mutation bias observed in bacteria. This conclusion is corroborated by the mutation patterns derived from single-nucleotide polymorphisms (SNPs) of a strictly anaerobic bacteria-*Clostridium botulinum*. Although ROS production is precluded in this bacteria, it still shows strong G:C → A:T mutation bias (Fig. 1A).

Besides ROS damage, UV-radiation also leads to DNA mutation by inducing the formation of dipyrimidines and the deamination of cytosines [13]. In fact, this mechanism has been used by Ossow-

* Corresponding author at: National Key Laboratory of Crop Genetic Improvement, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China. Fax: +86 27 87280877.
E-mail address: zhy630@mail.hzau.edu.cn (H.-Y. Zhang).

0006-291X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2011.05.017

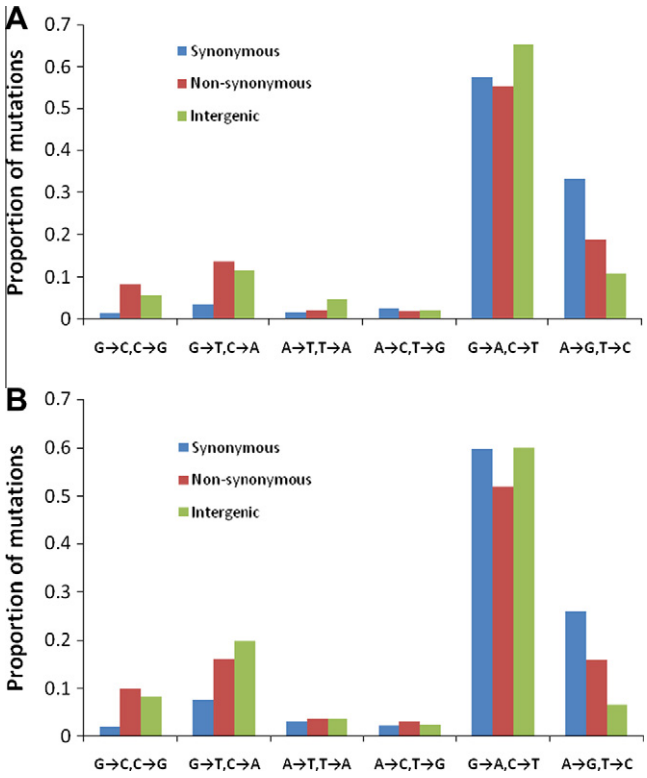

Fig. 1. Relative rates of the six nucleotide pair mutations in *Clostridium botulinum* (A) and *Sulfolobus islandicus* (B) genomes normalized by initial GC or AT contents. It can be seen that G:C → A:T mutations are more prevalent than A:T → G:C counterparts in various regions ($P < 0.0001$, *t*-test, random sampling 100 times, 10% mutation sites per time). Since *Clostridium botulinum* are strictly anaerobic bacteria, ROS damage is unlikely responsible for the G:C → A:T mutation bias of the bacteria. The SNP data for genomes of *Clostridium botulinum* (including eight closely related strains) and *Sulfolobus islandicus* (including seven closely related strains) (Tables S4 and S5) were downloaded from NCBI Entrez Genome Project database (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). The multiple genome alignment was performed by progressiveMauve with default parameters [29]. The criterion for selecting SNP-containing genes is that long completely matched sequences ($\geqslant$50 bp) exist in the upstream and downstream flanking regions of the SNP site. The mutation directions of the SNPs were determined by the routine methodology, that is, the abundant sites are regarded as ancestor states while the rare sites (occurring only once) are mutations [30].

ski and co-workers to elucidate the G:C → A:T mutation bias in *Arabidopsis thaliana* [7]. It is of great interest to examine whether the mutation bias in the 149 bacteria comes from UV-radiation. Through examining the habitat information recorded in NCBI Entrez Genome Project database (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi), it was found that a large part (~65%) of the bacteria that have a positive Z–Z_{pred} are animal parasites (Table S1). Since animal parasites are scarcely exposed to UV-radiation, UV-radiation is not likely responsible for the mutation bias in these bacteria. Indeed, there is no significant difference between the average Z–Z_{pred} values of animal parasites and of other bacteria ($P = 0.2973$, Wilcoxon two-sample test) (Table S3). In addition, UV-radiation cannot explain the mutation bias observed in humans, because the human germ cells are shielded from UV-radiation by the human body. Thus, other mechanisms should be considered to explain the universal mutation bias.

It is well known that CpG hypermutability leads to C → T mutations through spontaneous deamination of 5-methylcytosine [14], which provides a possible explanation to the G:C → A:T mutation bias. In the coding regions of some plant and animal genomes, CpG dinucleotides are frequently methylated at sites of C, which indeed make meaningful contributions to G:C → A:T transition bias in vertebrates and plants [7,10]. However, CpG hypermutability is unlikely a common basis for the universal mutation bias, because the mutation bias also occurs in the organisms that have no CpG methylation, such as yeast, *Drosophila melanogaster* and *C. elegans* [15]. Besides, the mutation patterns derived from SNPs of human genomes indicates that the G:C → A:T mutation bias still exists in various regions (3′UTR, 5′UTR, exons and introns), even if the CpG dinucleotides are ignored (Fig. 2).

In summary, it seems that the three well-known DNA mutation models are not sufficient to explain the universal DNA mutation bias. Thus, we should go beyond these models to find a common basis for the universal DNA mutation bias.

Forty years ago, Löwdin proposed a DNA mutation model, which was built on inter-base double proton transfers (DPT) [16]. This model argued that following the DPT, the four bases (A, G, C, T) become tautomeric forms (A*, G*, C*, T*), which do not bind to the normal Watson–Crick partners but to others, especially C, T, A, G, respectively. As a result, base transitions occur (Fig. 3). This model has been preliminarily supported by theoretical calculations [17,18]. The photoinduced DPT also has been observed experimentally in GC pairs, cytosine dimmers and model DNA base pairs

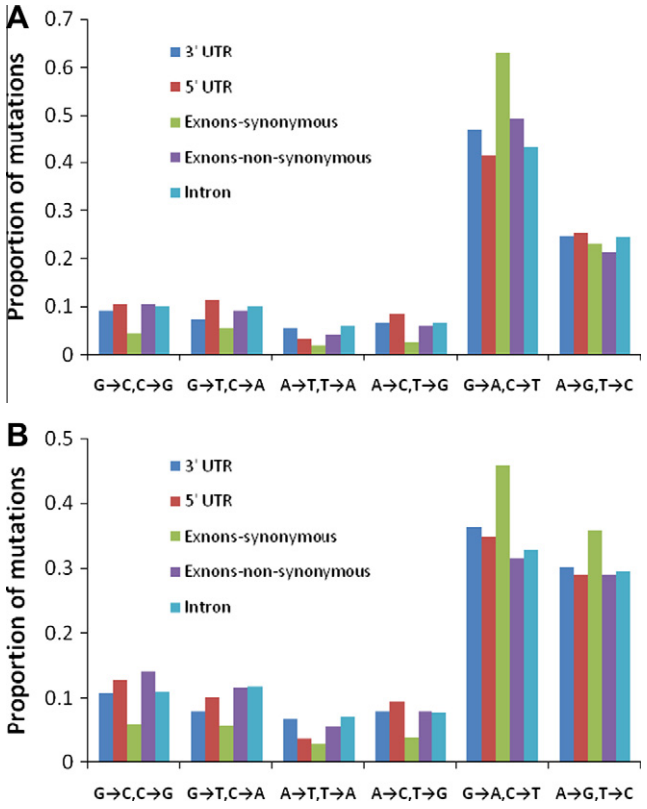

Fig. 2. Relative rates of the six nucleotide pair mutations in human genome normalized by initial GC or AT contents. (A) For total genome; (B) without CpG sites. It can be seen that G:C → A:T mutations are more prevalent than A:T → G:C counterparts in various regions of total genome ($P < 0.0001$, *t*-test, random sampling 100 times, 10% mutation sites per time). If CpG sites are deleted, the G:C → A:T mutation bias still exists in 3′UTR, 5′UTR, exons and introns ($P < 0.0001$, *t*-test, random sampling 100 times, 10% mutation sites per time). This suggests that CpG hypermutation is not the single reason for the G:C → A:T mutation bias in human genomes. The SNP data for Han Chinese in Beijing, China were downloaded from International HapMap Project web site (http://hapmap.ncbi.nlm.nih.gov) and NCBI database of single nucleotide polymorphisms (dbSNP) (http://www.ncbi.nlm.nih.gov/snp) (Tables S6 and S7). The criteria for SNP selection are: (i) minor allele frequency $\geqslant$0.05; (ii) derived allele frequency $\leqslant$0.1. The mutation directions of the SNPs were determined according to the information provided by NCBI (http://www.ncbi.nlm.nih.gov/snp). During the calculation of substitution numbers, both the mutations at transcribed and non-transcribed strands were considered [4].

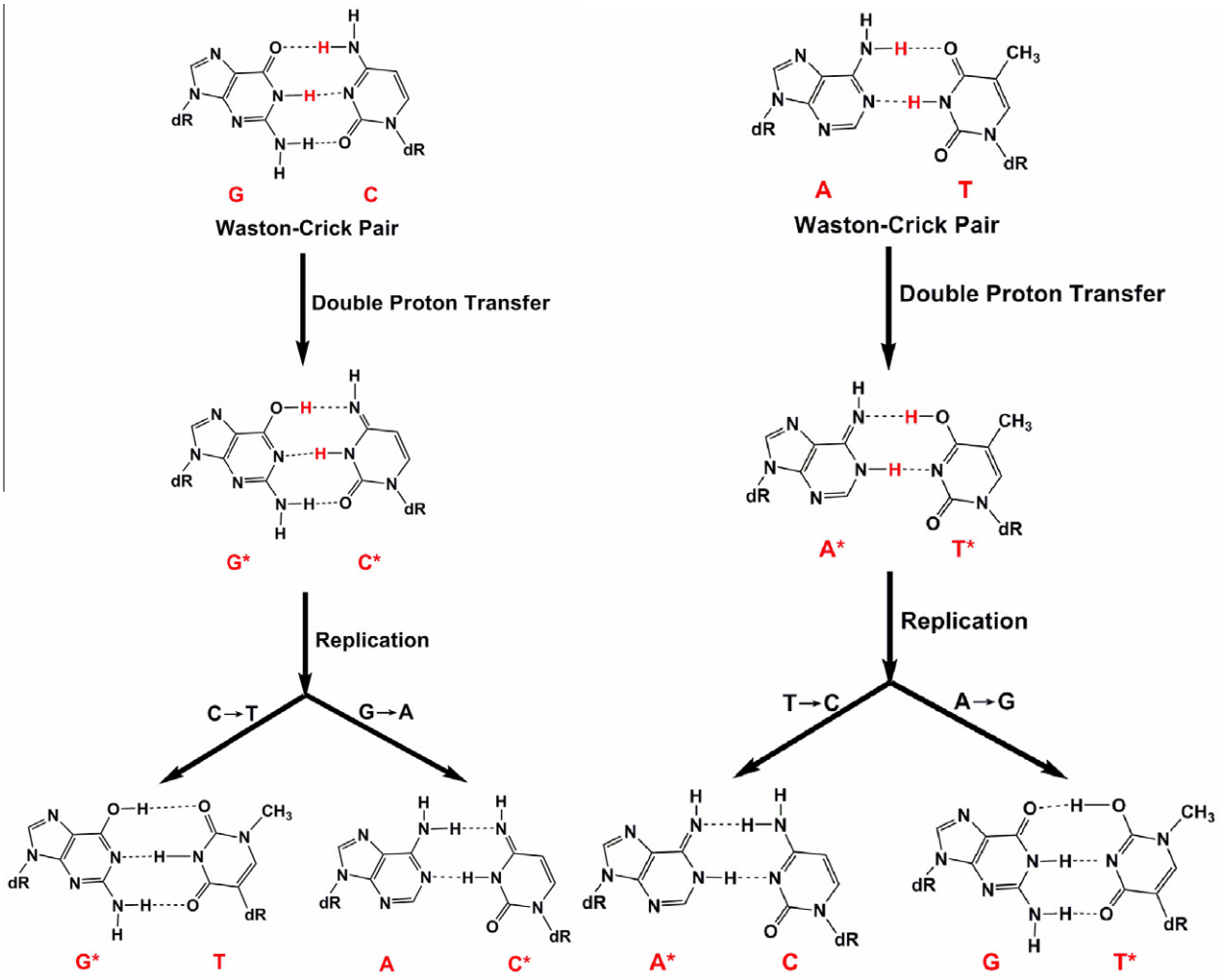

Fig. 3. Löwdin DNA mutation model based on inter-base double proton transfers (DPT). dR, deoxyribose. Since DPT is a fundamental and spontaneous chemical process and occurs more frequently within GC pairs than AT pairs, this model provides a common explanation for the universal mutation bias observed in archaea, bacteria, fungi, plants and animals.

[19–21]. As the free energy difference between GC and G*C* pairs is significantly smaller than that between AT and A*T* pairs (9.0–11.14 *vs.* 12.9 ~16.6 kcal/mol), the DPT occurs more frequently within GC pairs than AT pairs (10^{-7}–10^{-9} *vs.* 10^{-9}–10^{-12} per base pair) [17,18], which implies that the G:C → A:T mutations are more prevalent than the converse mutations. Because DPT is a spontaneous chemical process and exists in the DNA of all organisms, it is very appropriate to serve as the common molecular basis for the universal G:C → A:T mutation bias. This explanation also suggests that the biased mutation is inherent to the three domains of life and thus can be observed in archaea. A preliminary SNP analysis indeed reveals such kind of DNA mutation bias in archaea-*Sulfolobus islandicus* (Fig. 1B).

Since DNA mutation is indispensable to all organisms, Löwdin mutation model has broad biological implications. First, it is helpful to understand why G + C contents vary so dramatically in bacterial genomes (ranging from 16.5% to 75%) [8,9] and in coding and non-coding regions of vertebrate genomes [22]. Löwdin model implies that there exists an intrinsic G:C → A:T biased mutation in DNA of all organisms, which means that the genome compositions will get more and more AT-rich under relaxed selection, until a balance is reached between G:C → A:T and A:T → G:C mutations [8]. Therefore, the dramatic variations in G + C content of bacterial genomes are more likely to result from adaptive evolution rather than from different mutation patterns of bacterial genomes, as suggested by early studies [23,24]. Besides, as coding regions are under stronger selective pressure than non-coding regions [25], the latter are more possible to retain the AT-biased mutations than the former, which is helpful to explain why coding regions are more GC-rich than non-coding counterparts in vertebrate genomes [22].

Second, Löwdin model is helpful to understand the G/A bias of retroid viruses. The host-induced guanine-to-adenine (G → A) hypermutation plays an important role in the defense against retroid viruses [26,27]. It is thus of great interest to explore the mechanisms underlying the G/A bias in these viruses. Recently, Müller and Bonhoeffer argued that this bias is probably not a result of host-induced mutational pressure, but rather reflects a general predisposition associated with reverse transcription [28]. However, the molecular basis for this predisposition remains to be elucidated [28]. It is well-known that during the infection of retroid viruses, the RNA genome of the viruses will be copied and inserted into the host nuclear genome, which means that the virus genome will behave like host genome to undergo a G:C → A:T biased mutation in both strands. This is very likely to be the reason underlying the G/A bias of retroid viruses. This explanation is supported by the fact that although the retroviral genome is a single strand, the G → A and C → T silent nucleotide bias are strongly correlated, suggesting that the biased mutations occur in both strands [28]. Since the deletion of CpG sites and dipyrimidine sites has no impact on

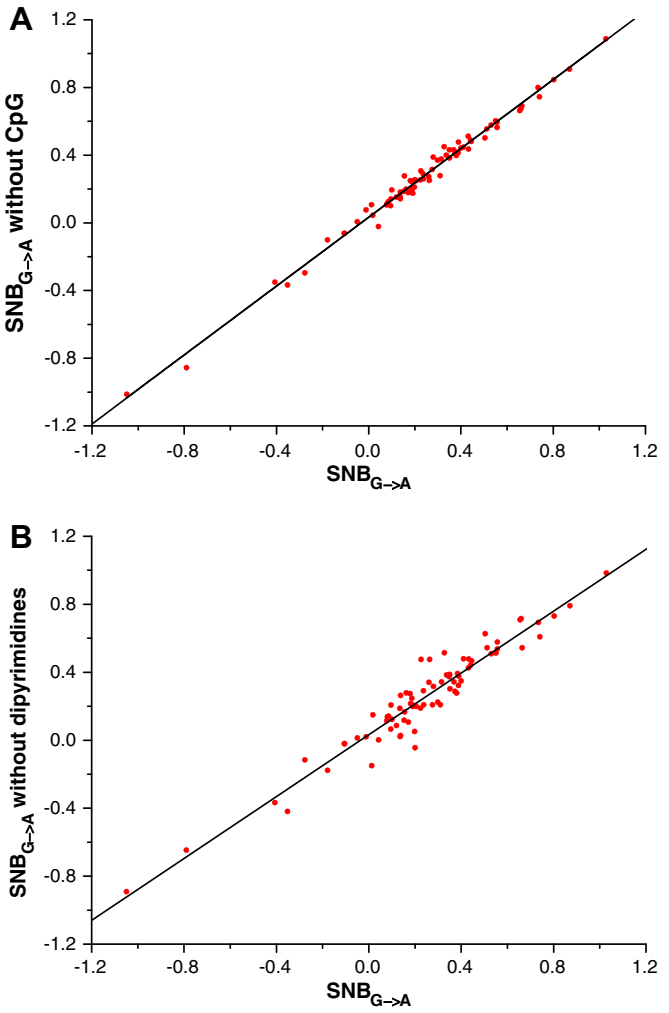

Fig. 4. Strong correlation between G → A silent nucleotide bias ($SNB_{G\to A}$) with and without CpG sites (A) and dipyrimidines (B) in retroviral genomes ($R > 0.96$, $P < 0.0001$, $n = 78$), indicating that deletion of CpG sites and dipyrimidines has no impact on the G → A silent nucleotide bias. The definition and calculation of $SNB_{G\to A}$ can refer to Ref. [28]. The original data were provided by Dr. Viktor Müller.

the G → A silent nucleotide bias in the viruses (Fig. 4), the CpG hypermutability and UV-radiation are unlikely responsible for the G/A bias of retroid viruses. Thus, DPT-based DNA mutation is a very appropriate model to explain the G → A silent nucleotide bias in retroid viruses. This explanation implies that the G/A bias of retroid viruses is a result of general predisposition associated with host (including animals and plants) mutation bias.

In conclusion, more than 40 years have passed since Löwdin proposed his DPT-based DNA mutation model. This model has attracted wide attention from chemists and physicists, but was largely overlooked by biologists. In this paper, we show the power of this model in explaining some intriguing observations in genetics and evolutionary biology. Since this model does not conflict with other DNA mutation models, we argue that it is very appropriate to serve as a basic framework to establish a more accurate theory for mutation mechanisms and thus deserves a great deal of attention.

Acknowledgments

We are grateful to Dr. Viktor Müller for sharing the original data of retroid virus genomes. This study was supported by the National Basic Research Program of China (Grant 2010CB126100), the National Natural Science Foundation of China (Grant 30870520), the China National Fundamental Fund of Personnel Training and the Fundamental Research Funds for the Central Universities.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2011.05.017.

References

[1] P.R. Haddrill, B. Charlesworth, Non-neutral processes drive the nucleotide composition of non-coding sequences in *Drosophila*, Biol. Lett. 4 (2008) 438–441.
[2] P.A. Lind, D.I. Andersson, Whole-genome mutational biases in bacteria, Proc. Natl. Acad. Sci. USA 105 (2008) 17878–17883.
[3] M. Lynch, W. Sung, K. Morris, N. Coffey, C.R. Landry, E.B. Dopman, W.J. Dickinson, K. Okamoto, S. Kulkarni, D.L. Hartl, W.K. Thomas, A genome-wide view of the spectrum of spontaneous mutations in yeast, Proc. Natl. Acad. Sci. USA 105 (2008) 9272–9277.
[4] G. Albrecht-Buehler, The spectra of point mutations in vertebrate genomes, Bioessays 31 (2009) 98–106.
[5] K.J. Balbi, E.P. Rocha, E.J. Feil, The temporal dynamics of slightly deleterious mutations in *Escherichia coli* and *Shigella* spp., Mol. Biol. Evol. 26 (2009) 345–355.
[6] D.R. Denver, P.C. Dolan, L.J. Wilhelm, W. Sung, J.I. Lucas-Lledo, D.K. Howe, S.C. Lewis, K. Okamoto, W.K. Thomas, M. Lynch, C.F. Baer, A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes, Proc. Natl. Acad. Sci. USA 106 (2009) 16310–16314.
[7] S. Ossowski, K. Schneeberger, J.I. Lucas-Lledo, N. Warthmann, R.M. Clark, R.G. Shaw, D. Weigel, M. Lynch, The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*, Science 327 (2010) 92–94.
[8] R. Hershberg, D.A. Petrov, Evidence that mutation is universally biased towards AT in bacteria, PLoS Genet. 6 (2010) e1001115.
[9] F. Hildebrand, A. Meyer, A. Eyre-Walker, Evidence of selection upon genomic GC-content in bacteria, PLoS Genet. 6 (2010) e1001107.
[10] M. Lynch, Rate, molecular spectrum, and consequences of human mutation, Proc. Natl. Acad. Sci. USA 107 (2010) 961–968.
[11] K.C. Cheng, D.S. Cahill, H. Kasai, S. NishimuraL, A. Loeb, 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G → T and A → C substitutions, J. Biol. Chem. 267 (1992) 166–172.
[12] V. Thiviyanathan, A. Somasunderam, D.E. Volk, T.K. Hazra, S. Mitra, D.G. Gorenstein, Base-pairing properties of the oxidized cytosine derivative, 5-hydroxy uracil, Biochem. Biophys. Res. Commun. 366 (2008) 752–757.
[13] Y. Barak, O. Cohen-Fix, Z. Livneh, Deamination of cytosine-containing pyrimidine photodimers in UV-irradiated DNA. Significance for UV light mutagenesis, J. Biol. Chem. 270 (1995) 24174–24179.
[14] P. Bird, DNA methylation and the frequency of CpG in animal DNA, Nucleic Acids Res. 8 (1980) 1499–1504.
[15] A. Jeltsch, Molecular biology. Phylogeny of methylomes, Science 328 (2010) 837–838.
[16] P.O. Löwdin, Proton tunneling in DNA and its biological implications, Rev. Mod. Phys. 35 (1963) 724–732.
[17] J. Florián, J. Leszczyński, Spontaneous DNA mutations induced by proton transfer in the guanine–cytosine base pairs: an energetic perspective, J. Am. Chem. Soc. 118 (1996) 3010–3017.
[18] G. Villani, Theoretical investigation of hydrogen atom transfer in the adenine–thymine base pair and its coupling with the electronic rearrangement. Concerted vs. stepwise mechanism, Chem. Phys. 12 (2010) 2664–2669.
[19] E. Nir, K. Kleinermanns, M.S. de Vries, Pairing of isolated nucleic-acid bases in the absence of the DNA backbone, Nature 408 (2000) 949–951.
[20] L. Blancafort, J. Bertran, M. Sodupe, Triplet (pi,pi) reactivity of the guanine–cytosine DNA base pair: benign deactivation versus double tautomerization via intermolecular hydrogen transfer, J. Am. Chem. Soc. 126 (2004) 12770–12771.
[21] O.-H. Kwon, A.H. Zewail, Double proton transfer dynamics of model DNA base pairs in the condensed phase, Proc. Natl. Acad. Sci. USA 104 (2007) 8703–8708.
[22] G. Bernardi, Isochores and the evolutionary genomics of vertebrates, Gene 241 (2000) 3–17.
[23] E. Freese, On the evolution of the base composition of DNA, J. Theor. Biol. 3 (1962) 82–101.
[24] N. Sueoka, On the genetic basis of variation and heterogeneity of DNA base composition, Proc. Natl. Acad. Sci. USA 48 (1962) 582–592.
[25] N. Jareborg, E. Birney, R. Durbin, Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs, Genome Res. 9 (1999) 815–824.
[26] R.S. Harris, K.N. Bishop, A.M. Sheehy, H.M. Craig, S.K. Petersen-Mahrt, I.N. Watt, M.S. Neuberger, M.H. Malim, DNA deamination mediates innate immunity to retroviral infection, Cell 113 (2003) 803–809.
[27] B. Mangeat, P. Turelli, G. Caron, M. Friedli, L. Perrin, D. Trono, Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts, Nature 424 (2003) 99–103.

[28] V. Müller, S. Bonhoeffer, Guanine–adenine bias: a general property of retroid viruses that is unrelated to host-induced hypermutation, Trends Genet. 21 (2005) 264–268.
[29] A.E. Darling, B. Mau, N.T. Perna, progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement, PLoS One 5 (2010) e11147.
[30] K.E. Holt, J. Parkhill, C.J. Mazzoni, P. Roumagnac, F.X. Weill, I. Goodhead, R. Rance, S. Baker, D.J. Maskell, J. Wain, C. Dolecek, M. Achtman, G. Dougan, High-throughput sequencing provides insights into genome variation and evolution in *Salmonella typhi*, Nat. Genet. 40 (2008) 987–993.