NRC Research Press

ARTICLE

The evolution of genomic GC content undergoes a rapid reversal within the genus *Plasmodium*

Hamid Nikbakht, Xuhua Xia, and Donal A. Hickey

Abstract: The genome of the malarial parasite *Plasmodium falciparum* is extremely AT rich. This bias toward a low GC content is a characteristic of several, but not all, species within the genus *Plasmodium*. We compared 4283 orthologous pairs of protein-coding sequences between *Plasmodium falciparum* and the less AT-biased *Plasmodium vivax*. Our results indicate that the common ancestor of these two species was also extremely AT rich. This means that, although there was a strong bias toward A+T during the early evolution of the ancestral *Plasmodium* lineage, there was a subsequent reversal of this trend during the more recent evolution of some species, such as *P. vivax*. Moreover, we show that not only is the *P. vivax* genome losing its AT richness, it is actually gaining a very significant degree of GC richness. This example illustrates the potential volatility of nucleotide content during the course of molecular evolution. Such reversible fluxes in nucleotide content within lineages could have important implications for phylogenetic reconstruction based on molecular sequence data.

Key words: malaria, nucleotide bias, biased gene conversion, nucleotide content, molecular phylogeny.

Résumé : Le génome du parasite responsable de la transmission du paludisme, *Plasmodium falciparum*, est extrêmement riche en AT. Ce biais en direction d'un faible contenu en GC est caractéristique de plusieurs mais pas toutes les espèces au sein du genre *Plasmodium*. Les auteurs ont comparé 4283 paires orthologues de séquences codantes entre le *Plasmodium falciparum* et le *Plasmodium vivax*, une espèce qui présente un biais plus faible. Les résultats indiquent que l'ancêtre commun de ces deux espèces était également très riche en AT. Cela suggère que, bien qu'il ait existé un fort biais en faveur des A+T au cours de l'évolution antérieure de l'ancêtre des *Plasmodium*, il se serait produit un renversement subséquent de cette tendance au cours de l'évolution plus récente chez certaines espèces comme le *P. vivax*. De plus, les auteurs montrent que le génome du *P. vivax* a non seulement perdu sa richesse en AT, mais encore il a acquis une richesse très significative en GC. Cet exemple illustre la volatilité potentielle de la composition en nucléotides au cours de l'évolution moléculaire. De tels changements réversibles en composition nucléotidique au sein d'un embranchement pourraient avoir des implications importantes pour la reconstruction de phylogénies sur la base des séquences moléculaires. [Traduit par la Rédaction]

Mots-clés : paludisme, biais nucléotidique, conversion génique biaisée, composition nucléotidique, phylogénie moléculaire.

Introduction

Genomic nucleotide content (usually measured as GC content) varies widely among prokaryotic genomes (Jukes and Bhushan 1986; Muto and Osawa 1987; Sueoka 1988; Haywood-Farmer and Otto 2003; Xia 2003) and mitochondrial genomes (Jermiin et al. 1994; Foster et al. 1997). In contrast to the prokaryote and organelle genomes, however, overall nucleotide content varies relatively little among the nuclear genomes of vertebrates, although there are significant variations in GC content within individual vertebrate genomes (Ikemura and Aota 1988; Galtier and Mouchiroud 1998). But if we consider other nonvertebrate eukaryotic groups, e.g., insects, plants, and protists, we can also see considerable levels of variation in GC content between genomes (see, e.g., Serres-Giardi et al. 2012). In this paper, we consider the nucleotide content of the malarial parasite *Plasmodium falciparum*, which is known to have one of the lowest values for genomic GC content among eukaryotes (Gardner et al. 2002).

The intergenomic variation in nucleotide content is correlated with a corresponding bias in the amino acid composition of the encoded proteomes (Collins and Jukes 1993; Foster et al. 1997; Singer and Hickey 2000; Wang et al. 2004; Bastien et al. 2004; Urbina et al. 2006). The underlying causes of these variations in nucleotide content are of considerable interest to molecular evolutionary biologists. The main focus of interest has been on whether these variations are a reflection of the action of natural selection (see, e.g., Xia 1996; Xia and Palidwor 2005) or whether they reflect some form of biased mutational pressure (Sueoka 1988; Wang et al. 1999, 2004; Galtier et al. 2001; Duret et al. 2006; Pessia et al. 2012; Lartillot 2013).

These correlated biases in nucleotide and amino acid content are also of great practical importance to evolutionary biologists because the accuracy of phylogenetic reconstruction based on molecular sequence data can be misleading if such biases are not properly taken into account (Foster and Hickey 1999). Indeed, the unusually low GC content of the *Plasmodium* genome has presented unique challenges both for gene discovery (Gardner et al. 2002; Paila et al. 2008) and for phylogenetic reconstruction (Nabholz et al. 2011; Blanquart and Gascuel 2011).

Our motivation for this study was the observation that although *P. falciparum* has a very AT-rich genome, less than 20% GC, other species within the genus *Plasmodium*, such as *P. vivax*, are characterized by genomic GC contents that are much closer to 50% GC

Received 21 October 2014. Accepted 13 December 2014.

Corresponding Editor: B. Golding.

H. Nikbakht and D.A. Hickey. Department of Biology, Concordia University, Montreal, QC H4B 1R6, Canada.
X. Xia. Department of Biology, University of Ottawa, Ottawa, ON K1N 6N5, Canada.
Corresponding author: Donal A. Hickey (e-mail: donal.hickey@concordia.ca).

(Carlton et al. 2008). We wished to know if the common ancestor of these two species had an AT-biased genome like that of *P. falciparum* or a relatively unbiased genome like that of *P. vivax*. Our results support the first option: i.e., they indicate that the common ancestor of these two species was characterized by an extremely low GC content like that of *P. falciparum*. Therefore, we can infer that although the ancestral lineage of the genus *Plasmodium* evolved toward an exceptionally low GC content, some species within the genus, e.g., *P. vivax*, have subsequently evolved in the opposite direction, i.e., toward higher GC content. The implication of this result is that the evolutionary changes in genomic nucleotide content can be much more rapid and more complex than is generally appreciated.

Methods

We used the PlasmoDB database, release 7.1 (Bahl et al. 2003), to access the complete genome sequences for six species of *Plasmodium*: *P.vivax*, *P. falciparum*, *P. knowlesi*, *P. berghei*, *P. chabaudi*, and *P. yoelii*. Since these genomes were fully annotated we were able to identify and download the coding sequences for all of the protein-coding genes. We focussed on the protein-coding regions because this allowed us to align homologous sequences. Next, we calculated the average nucleotide content of the coding sequences for each of the six species.

Based on the average GC contents of the six species, we chose to focus on the comparison of two species, *P. falciparum* and *P. vivax*. We calculated the GC content of the coding sequences as follows. We began with 5491 coding sequences from *P. falciparum* and 5435 coding sequences from *P. vivax* but, after discarding unusually long (greater than 30 kb) and short (less than 150 bp) sequences, along with sequences lacking an ATG start codon, we retained 5449 coding sequences from *P. falciparum* and 5335 from *P. vivax*. We also downloaded the amino acid sequences corresponding to these coding regions for the next steps in our analysis.

Next, we identified orthologous sequences between these two species and calculated the GC contents in orthologous pairs of genes. To create the orthologous datasets, we aligned the 5449 *P. falciparum* protein sequences against 5335 protein sequences from *P. vivax* using standalone BLAST (Altschul et al. 1990). We used protein sequences for the alignment to avoid any bias arising from codon degeneracy affecting the homology search. Those hits with expectation values larger than $1E^{-20}$ and low alignment scores were discarded. The resulting dataset contains 4283 orthologous gene pairs from *P. falciparum* and *P. vivax*.

To calculate the GC content at conserved and variable sites, we performed pairwise alignments, using stand-alone ClustalW (Larkin et al. 2007). We began by aligning each pair of orthologous protein sequences (4283 orthologous pairs). Next, we removed the gapped columns and trimmed the unaligned end sequences. We then assigned the nucleotide sequences to their corresponding protein sequences and removed the codons corresponding to amino acid gaps in the corresponding protein alignment. We used this dataset for all the analyses on conserved and variable sites, as well as different codon positions.

Using the 4283 DNA alignments, we first calculated the GC content of the individual orthologs (shown in the supplementary data, Table S1[1]). We then partitioned the data into conserved sites and variable sites as follows: conserved sites are those sites in the alignment at which the same nucleotide is found in both sequences, while variable sites are those sites where there is nucleotide difference between the two aligned sequences. We then calculated the GC content of the conserved and variable sites separately. Finally, we partitioned the data by codon position and calculated the GC content at each of the three codon positions separately (see Table S2).

Table 1. GC content of six species of *Plasmodium*.

Species	G+C content (%)
Plasmodium vivax	46.2
Plasmodium falciparum	23.8
Plasmodium knowlesi	40.2
Plasmodium berghei	23.7
Plasmodium chabaudi	25.5
Plasmodium yoelii	24.2

Note: These values are based on the GC content of the coding sequences only (i.e., introns and intergenic noncoding sequences are not included). Note that the exclusion of noncoding sites results in values that are a little higher than the results based on whole genome sequences.

Results

We first compared the average GC content of the protein-coding sequences from six different species within the genus *Plasmodium* that were available on the *Plasmodium* database (http://www.plasmodb.org). The results are shown in Table 1. From the Table, we can see that four species, including *P. falciparum*, have very low GC contents, while two species, *P. knowlesi* and *P. vivax*, have relatively high GC contents. We chose to focus our attention on the two species *P. falciparum* and *P. vivax* because both species are important human parasites and because their coding sequences differ in GC content by almost a factor of two (see Table 1).

Our goal was to trace the patterns of nucleotide substitution since the divergence of the GC-poor *P. falciparum* and the seemingly GC-neutral *P. vivax*. Normally, one would use a phylogenetic approach to investigate such a problem. Specifically, one would plot the various species of *Plasmodium* on a phylogenetic tree and add some appropriate outgroup species. But this normal approach is complicated by some special challenges in the case of *Plasmodium*. Because of its lack of a fossil record, combined with its reduced morphology (a result of the parasitic lifestyle), most phylogenies for this group are based almost solely on molecular sequence data. But, as has been shown in a series of recent publications, these phylogenies are themselves prone to the effects of the nucleotide biases (e.g., Silva et al. 2011; Perkins 2014; Bensch et al. 2013). In other words, the phylogenetic reconstruction methods presuppose a prior knowledge of the effect that we are trying to investigate. Ideally, one would like to use a phylogeny-free approach.

By restricting ourselves to some very basic questions about the evolution of nucleotide content since the divergence of these two species, we have been able to avoid invoking a particular phylogenetic branching pattern. Basically, we posed some simple questions about the general statistical properties of the sequences rather than asking about particular patterns of nucleotide substitution. We were inspired by the simple approach used by M. Kimura to show that rates of substitution at synonymous sites were more frequent than at nonsynonymous sites. This was achieved by a simple comparison of aligned sequence pairs (Kimura 1981).

As mentioned in the Methods, we identified 4283 orthologous coding sequences between *P. falciparum* and *P. vivax*. We then calculated the GC content of each of the orthologs. In addition to calculating the overall GC content of each gene, we also identified the conserved and nonconserved nucleotides in the alignments of each orthologous pair and calculated the GC content of each type of site (i.e., conserved sites and variable sites). We reasoned that,

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/gen-2014-0158.

on average, the conserved sites would tend to reflect the ancestral sequence composition, whereas the nonconserved sites would provide an indication of the changed sequence composition since the divergence of the two species. Our results (see Table 1; Fig. 1) show that the conserved sites in the apparently GC-neutral *P. vivax* are very AT rich (GC content 27.63% ± 0.08%), whereas the variable sites are, in fact, strongly GC biased (GC content 72.08% ± 0.10%). In other words, the seemingly unbiased overall sequence of *P. vivax* (overall GC content 45.7% ± 0.08% for the orthologous subset of genes) is a result of averaging a strong AT bias at the conserved sites and an equally strong GC bias at the variable sites. Since this pattern is seen consistently in all of the 4283 aligned sequence pairs (see Table S1), the effects are highly significant ($p < 0.0001$).

In an effort to explore this large difference in GC content between the conserved and variable sites, we decided to examine the *P. vivax* coding sequences more closely. To do this, we calculated the GC content of these sequences at each of the three codon positions separately. This was done because the effects of nucleotide bias are often seen most clearly at the third codon position, while there is usually least bias seen at the selectively constrained second codon position. The results, shown in Fig. 2 and in Table S2, reveal a complex pattern. We see a strong AT bias at the second codon position (GC content 32.95% ± 0.09%) and a significant GC bias at the third codon position (GC content 58.60% ± 0.14%); the first codon position has an intermediate nucleotide composition (GC content 45.55% ± 0.09%). Moreover, the effect is again very consistent over the 4283 individual genes (see Table S2), which again results in a high level of statistical significance for these differences between codon positions ($p < 0.0001$).

Taken separately, the two different patterns of nucleotide content heterogeneity presented in Figs. 1 and 2 are difficult to explain. By combining the two results, however, we can come up with a relatively simple explanation (see Discussion below).

Discussion

Let us first consider the variations in nucleotide content at each of the three codon positions in *P. vivax* (see Fig. 2). This pattern defies a simple explanation based on either the action of natural selection or of mutational bias. A selectionist explanation would have to invoke translational selection at the third codon positions that favored a high GC content. In addition, we would have to invoke a simultaneous selection based on the coding capacity of the second codon position that happened to favor a low GC content. And, finally, we would have to explain why these complicated patterns of selection do not affect other species of *Plasmodium*, such as *P. falciparum*. A consideration of a simple mutational bias, including the mismatch repair biases associated with biased gene conversion, does not provide a plausible explanation either; we would have to invoke a mutational bias toward higher than average GC at the third codon position along with a simultaneous bias toward low GC at the neighboring second codon position.

If we now consider this same pattern of codon-specific heterogeneity in GC content in light of the information provided in Fig. 1, we can come up with a relatively simple explanation. Specifically, considering that the ancestor of *P. vivax* had an AT-rich genome at the point of its divergence from *P. falciparum*, all we need to invoke is an overall mutational bias toward high GC, but starting from this overall low GC level. It is already known that mutational bias affects the less constrained third codon position more than it affects the selectively constrained second codon position (Kimura 1981). Therefore, an overall mutational bias toward high GC, starting from an AT-rich ancestral sequence, and filtered through different levels of selective constraint at each of the three codon positions, would yield the pattern observed in Fig. 2. In other words, the data presented in Figs. 1 and 2 fit nicely together to provide us with an explanation of the current complex composition of the *P. vivax* sequences. According to this explanation, the

Fig. 1. GC content of *Plasmodium vivax* coding sequences. The left bar indicates GC content of the sites that are conserved in the alignment with the orthologous sequence from *Plasmodium falciparum*, while the middle bar represents the GC content of the variable, i.e., nonconserved sites; the right-hand bar indicates the GC content of the entire *P. vivax* sequence (conserved and variable sites combined). The data are the averages of 4283 individual genes. Because the number of gene sequences was large and the trend was very consistent among the sequences, the standard errors (values given in the Results) were of the order of 0.1% or less; such values are too small to be visible in the Figure. The dotted line indicates 50% GC (i.e., the unbiased value). The raw data for the 4283 individual genes are shown in the supplementary data, Table S1.

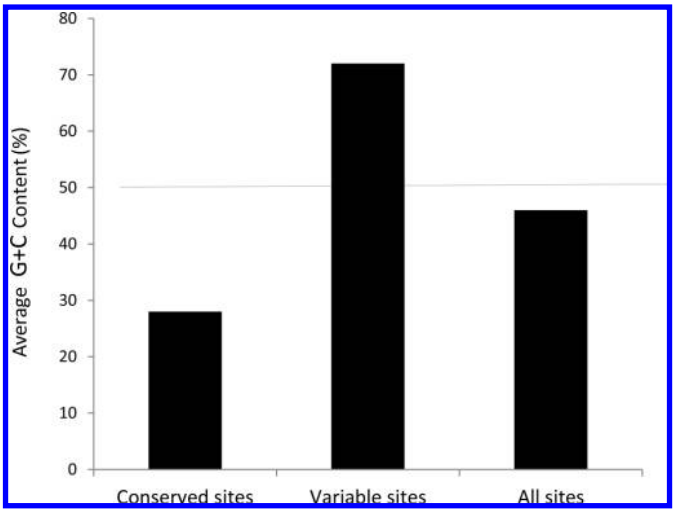
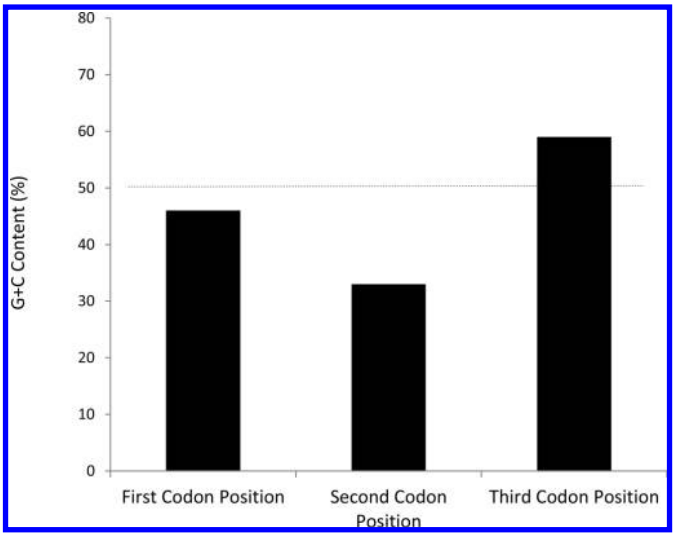


Fig. 2. The average GC content of *Plasmodium vivax* genes at each of the three codon positions. These values were also based on the averages of the 4283 individual genes. The raw data for the individual genes are shown in the supplementary data, Table S2.



P. vivax genome, while appearing to be relatively GC neutral overall, is actually in transition from an ancestral AT-rich genome toward becoming a very GC-rich genome. For example, the GC content at the third codon position among the variable sites is 76% GC. This illustrates how quickly (in macroevolutionary terms) a genome may switch from AT richness to GC richness, and it also raises a number of unanswered questions.

The most obvious question is why there should be a relatively rapid increase in the genomic GC content of *P. vivax* since its

divergence from its AT-rich ancestor, especially given that there is no such increase seen within the *P. falciparum* lineage. While we cannot give a definitive answer to this question, a number of suggestions can be made based on what we already know about the mutational biases and DNA repair biases. Early studies on mismatch repair within artificially created DNA heteroduplexes showed that the AT bias of mutation is balanced by a GC bias in DNA heteroduplex repair (Brown and Jiricny 1988). At about the same time (Bishop et al. 1987) it was shown that meiotic gene conversion in yeast results from the correction of heteroduplex DNA. Combining these two results led to the prediction that increased levels of recombination and gene conversion could lead to increases in GC content because of the increased frequency of biased heteroduplex repair. Support from this prediction came from the observation that duplicated genes undergoing concerted evolution within the *Drosophila* genome become GC rich because they are subjected to increased levels of recombination and, consequently, increased levels of biased heteroduplex repair (Hickey et al. 1991; Wang et al. 1999; Bettencourt and Feder 2002). This idea was applied more generally to the GC-content evolution of mammalian genomes by Galtier et al. (2001) who referred to the phenomenon as biased gene conversion. Duret et al. (2006) suggested that such GC-biased heteroduplex repair associated with meiotic recombination could explain variations in GC content within human genomes. Smith and Lee (2008) invoked a similar explanation for differences in GC content between mitochondrial genomes. Recently, there is growing evidence that this mechanism is widespread among eukaryotes (Pessia et al. 2012) and that it is a major determinant of the nucleotide composition of mammalian genomes (Lartillot 2013).

The studies (cited above) of duplicated genes in the *Drosophila* genome show that the GC content can change even when the species-specific rates of mutation and repair remain the same; what changes in the case of the duplicated genes is the frequency of heteroduplex formation. Considering these facts, we have asked in what ways the genome of *P. vivax* might be subject to less AT-biased mutation, and (or) more GC-biased heteroduplex repair, i.e., more biased gene conversion, than that of *P. falciparum*. Some clues may come from considering differences in the biology of the two species. First, there is evidence that *P. vivax* infects reticulocyte cells preferentially, whereas *P. falciparum* infects red blood cells at all stages, including the oxygen-rich mature erythrocytes (see Gunalan et al. 2013 for a recent review). This might explain why the *P. vivax* genome would be subject to less oxidative damage and, consequently, less AT-biased mutation. But we need to explain not only that the *P. vivax* sequences are less AT biased, we also need to explain why they are becoming GC rich at the variable sites. This increase in GC suggests the possibility that the *P. vivax* sequences might be subject to increased levels of heteroduplex repair. It turns out that the population biology of *P. vivax* lends some support to this idea. Specifically, it has been shown that *P. vivax* is characterized by higher frequencies of multiple-clone infections than *P. falciparum* and that *P. vivax* also shows higher levels of sequence polymorphism (see, e.g., Havryliuk and Ferreira 2009). These population characteristics would result in a higher frequency of heteroduplex formation during meiotic recombination in *P. vivax*, as compared to the more monomorphic and inbred *P. falciparum*. A further piece of evidence that the increased GC content of the *P. vivax* nuclear genome is related to meiotic recombination comes from the fact that the mitochondrial sequences of this species (which are not subject to meiotic recombination and heteroduplex repair) remain extremely AT rich (McIntosh et al. 1998; Rathore et al. 2001; Blanquart and Gascuel 2011).

Regardless of the mechanisms underlying these changes in nucleotide content, our results have relevance to the construction of molecular phylogenies. For example, they alert us to the fact that the average nucleotide content of a group of organisms may not be a reliable guide to the patterns of nucleotide bias within each member of that group because the GC content of the group members can potentially oscillate over a wide range. The results also show, and perhaps more importantly, that a nucleotide content that appears to be GC neutral overall does not necessarily mean that there were no nucleotide biases occurring during the evolution of the lineage; specifically, the seemingly bias-free genome of *P. vivax* has been subject to a strong AT bias followed by a strong GC bias. Finally, our results illustrate the fact that the overall nucleotide content of a sequence may be a poor indicator of the biases at the variable sites, i.e., those sites where the evolution is actually happening; this would be especially true for recent divergences where the variable sites, being a small minority of all sites, would contribute relatively little to the overall average nucleotide composition.

In summary, our results indicate that the *P. vivax* genome is in the process of evolving relatively rapidly from an extremely AT-rich ancestral genome toward becoming a GC-rich genome. This shows that genomic nucleotide content may be subject to more rapid fluxes than was previously appreciated. It also raises questions about the evolutionary forces underlying such switches in nucleotide composition. And finally it points to new complexities that should be taken into account when constructing molecular phylogenies.

Acknowledgements

This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to D.A.H., and by a Graduate Scholarship from Concordia University to H.N.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2. PMID:2231712.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M.J., Gajria, B., Grant, G.R., et al. 2003. PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. Nucleic Acids Res. **31**: 212–215. PMID:12519984.

Bastien, O., Lespinats, S., Roy, S., Métayer, K., Fertil, B., Codani, J.J., and Maréchal, E. 2004. Analysis of the compositional biases in *Plasmodium falciparum* genome and proteome using *Arabidopsis thaliana* as a reference. Gene, **336**: 163–173. doi:10.1016/j.gene.2004.04.029. PMID:15246528.

Bensch, S., Hellgren, O., Križanauskienė, A., Palinauskas, V., Valkiūnas, G., Outlaw, D., and Ricklefs, R.E. 2013. How can we determine the molecular clock of malaria parasites? Trends Parasitol. **29**: 363–369. doi:10.1016/j.pt.2013.03.011. PMID:23648374.

Bettencourt, B.R., and Feder, M.E. 2002. Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. J. Mol. Evol. **54**: 569–586. doi:10.1007/s00239-001-0044-7. PMID:11965431.

Bishop, D.K., Williamson, M.S., Fogel, S., and Kolodner, R.D. 1987. The role of heteroduplex correction in gene conversion in *Saccharomyces cerevisiae*. Nature, **328**: 362–364. doi:10.1038/328362a0. PMID:3299108.

Blanquart, S., and Gascuel, O. 2011. Mitochondrial genes support a common origin of rodent malaria parasites and *Plasmodium falciparum*'s relatives infecting great apes. BMC Evol. Biol. **11**: 70. doi:10.1186/1471-2148-11-70. PMID:21406081.

Brown, T.C., and Jiricny, J. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell, **54**: 705–711. doi:10.1016/S0092-8674(88)80015-1. PMID:2842064.

Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., et al. 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature, **455**: 757–763. doi:10.1038/nature07327. PMID:18843361.

Collins, D.W., and Jukes, T.H. 1993. Relationship between G + C in silent sites of codons and amino acid composition of human proteins. J. Mol. Evol. **36**: 201–213. doi:10.1007/BF00160475. PMID:8483158.

Duret, L., Eyre-Walker, A., and Galtier, N. 2006. A new perspective on isochore evolution. Gene, **385**: 71–74. doi:10.1016/j.gene.2006.04.030. PMID:16971063.

Foster, P.G., and Hickey, D.A. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. **48**: 284–290. doi:10.1007/PL00006471. PMID:10093217.

Foster, P.G., Jermiin, L.S., and Hickey, D.A. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. **44**: 282–288. doi:10.1007/PL00006145. PMID:9060394.

Galtier, N., and Mouchiroud, D. 1998. Isochore evolution in mammals: a human-like ancestral structure. Genetics, **150**: 1577–1584. PMID:9832533.

Galtier, N., Piganeau, G., Mouchiroud, D., and Duret, L. 2001. GC-content evolu-

tion in mammalian genomes: the biased gene conversion hypothesis. Genetics, **159**: 907–911. PMID:11693127.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature, **419**: 498–511. doi:10.1038/nature01097. PMID:12368864.

Gunalan, K., Gao, X., Yap, S.S., Huang, X., and Preiser, P.R. 2013. The role of the reticulocyte-binding-like protein homologues of *Plasmodium* in erythrocyte sensing and invasion. Cell Microbiol. **15**: 35–44. doi:10.1111/cmi.12038. PMID: 23046317.

Havryliuk, T., and Ferreira, M.U. 2009. A closer look at multiple-clone *Plasmodium vivax* infections: detection methods, prevalence and consequences. Mem. Inst. Oswaldo Cruz, **104**: 67–73. doi:10.1590/S0074-02762009000100011. PMID: 19274379.

Haywood-Farmer, E., and Otto, S.P. 2003. The evolution of genomic base composition in bacteria. Evolution, **57**: 1783–1792. doi:10.1554/01-535. 14503620.

Hickey, D.A., Bally-Cuif, L., Abukashawa, S., Payant, V., and Benkel, B.F. 1991. Concerted evolution of duplicated protein-coding genes in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. **88**: 1611–1615. doi:10.1073/pnas.88.5.1611. PMID:1900365.

Ikemura, T., and Aota, S. 1988. Global variation in G+C content along vertebrate genome DNA. Possible correlation with chromosome band structures. J. Mol. Biol. **203**: 1–13. PMID:3054117.

Jermiin, L.S., Graur, D., Lowe, R.M., and Crozier, R.H. 1994. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. J. Mol. Evol. **39**: 160–173. PMID:7932780.

Jukes, T.H., and Bhushan, V. 1986. Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. J. Mol. Evol. **24**: 39–44. doi:10.1007/BF02099949. PMID:3104617.

Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. U.S.A. **78**: 454–458. doi:10.1073/pnas.78.1.454. PMID:6165991.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics, **23**: 2947–2948.

Lartillot, N. 2013. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. Mol. Biol. Evol. **30**: 489–502. doi:10.1093/molbev/mss239. PMID:23079417.

McIntosh, M.T., Srivastava, R., and Vaidya, A.B. 1998. Divergent evolutionary constraints on mitochondrial and nuclear genomes of malaria parasites. Mol. Biochem. Parasitol. **95**: 69–80. doi:10.1016/S0166-6851(98)00093-0. PMID:9763290.

Muto, A., and Osawa, S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. U.S.A. **84**: 166–169. doi:10.1073/pnas.84.1.166. PMID:3467347.

Nabholz, B., Künstner, A., Wang, R., Jarvis, E.D., and Ellegren, H. 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. Mol. Biol. Evol. **28**: 2197–2210. doi:10.1093/molbev/msr047. PMID: 21393604.

Paila, U., Kondam, R., and Ranjan, A. 2008. Genome bias influences amino acid choices: analysis of amino acid substitution and re-compilation of substitution matrices exclusive to an AT-biased genome. Nucleic Acids Res. **36**: 6664–6675. doi:10.1093/nar/gkn635. PMID:18948281.

Perkins, S.L. 2014. Malaria's many mates: past, present and future of the systematics of the Order Haemosporida. J. Parasitol. **100**: 11–25. doi:10.1645/13-362.1. PMID:24059436.

Pessia, E., Popa, A., Mousset, S., Rezvoy, C., Duret, L., and Marais, G.A. 2012. Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol. Evol. **4**: 675–682. doi:10.1093/gbe/evs052. PMID:22628461.

Rathore, D., Wahl, A.M., Sullivan, M., and McCutchan, T.F. 2001. A phylogenetic comparison of gene trees constructed from plastid, mitochondrial and genomic DNA of *Plasmodium* species. Mol. Biochem. Parasitol. **114**: 89–94. doi:10.1016/S0166-6851(01)00241-9. PMID:11356517.

Serres-Giardi, L., Belkhir, K., David, J., and Glémin, S. 2012. Patterns and evolution of nucleotide landscapes in seed plants. Plant Cell, **24**: 1379–1397. doi: 10.1105/tpc.111.093674. PMID:22492812.

Silva, J.C., Egan, A., Friedman, R., Munro, J.B., Carlton, J.M., and Hughes, A.L. 2011. Genome sequences reveal divergence times of malaria parasite lineages. Parasitology, **138**: 1737–1749. doi:10.1017/S0031182010001575. PMID: 21118608.

Singer, G.A., and Hickey, D.A. 2000. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. Mol. Biol. Evol. **17**: 1581–1588. doi: 10.1093/oxfordjournals.molbev.a026257. PMID:11070046.

Smith, D.R., and Lee, R.W. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. Mol. Biol. Evol. **25**: 487–496. doi:10.1093/molbev/msm245. PMID:18222946.

Sueoka, N. 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. U.S.A. **85**: 2653–2657. doi:10.1073/pnas.85.8.2653. PMID:3357886.

Urbina, D., Tang, B., and Higgs, P.G. 2006. The response of amino acid frequencies to directional mutation pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. J. Mol. Evol. **62**: 340–361. doi:10.1007/s00239-005-0051-1. PMID:16477524.

Wang, H.C., Singer, G.A., and Hickey, D.A. 2004. Mutational bias affects protein evolution in flowering plants. Mol. Biol. Evol. **21**: 90–96. PMID:14595101.

Wang, S., Magoulas, C., and Hickey, D.A. 1999. Concerted evolution within a trypsin gene cluster in *Drosophila*. Mol. Biol. Evol. **16**: 1117–1124. doi:10.1093/oxfordjournals.molbev.a026202. PMID:10486967.

Xia, X. 1996. Maximizing transcription efficiency causes codon usage bias. Genetics, **144**: 1309–1320. PMID:8913770.

Xia, X. 2003. DNA methylation and *Mycoplasma* genomes. J. Mol. Evol. **57**(Suppl. 1): S21–S28. PMID:15008400.

Xia, X., and Palidwor, G. 2005. Genomic adaptation to acidic environment: evidence from *Helicobacter pylori*. Am. Nat. **166**: 776–784. doi:10.1086/497400. PMID:16475092.

This article has been cited by:

1. Emma R. Cold, Gerardo R. Vasta, José A. Fernández Robledo. 2017. Transient Expression of Plasmodium berghei MSP8 and HAP2 in the Marine Protozoan Parasite Perkinsus marinus. *Journal of Parasitology* **103**:1, 118-122. [CrossRef]
2. William L. Hamilton, Antoine Claessens, Thomas D. Otto, Mihir Kekre, Rick M. Fairhurst, Julian C. Rayner, Dominic Kwiatkowski. 2016. Extreme mutation bias and high AT content in Plasmodium falciparum. *Nucleic Acids Research* gkw1259. [CrossRef]
3. Yulong Wei, Juan Wang, Xuhua Xia. 2016. Coevolution between Stop Codon Usage and Release Factors in Bacterial Species. *Molecular Biology and Evolution* **33**:9, 2357-2367. [CrossRef]
4. Staffan Bensch, Björn Canbäck, Jeremy D. DeBarry, Tomas Johansson, Olof Hellgren, Jessica C. Kissinger, Vaidas Palinauskas, Elin Videvall, Gediminas Valkiūnas. 2016. The Genome of Haemoproteus tartakovskyi and Its Relationship to Human Malaria Parasites. *Genome Biology and Evolution* **8**:5, 1361-1373. [CrossRef]