J Mol Evol (2002) 55:260–264
DOI: 10.1007/s00239-002-2323-3

JOURNAL OF MOLECULAR EVOLUTION

Aerobiosis Increases the Genomic Guanine Plus Cytosine Content (GC%) in Prokaryotes

Hugo Naya,[1] Héctor Romero,[1,2] Alejandro Zavala,[1] Beatriz Alvarez,[3] Héctor Musto[1]

[1] Laboratorio de Organización y Evolución del Genoma, Facultad de Ciencias, Iguá 4225, Montevideo 11400, Uruguay
[2] Departamento de Genética, Facultad de Medicina, Gral. Flores 2125, Montevideo 11800, Uruguay
[3] Laboratorio de Enzimología, Facultad de Ciencias, Iguá 4225, Montevideo 11400, Uruguay

Received: 13 November 2001 / Accepted: 11 February 2002

Abstract. The huge variation in the genomic guanine plus cytosine content (GC%) among prokaryotes has been explained by two mutually exclusive hypotheses, namely, selectionist and neutralist. The former proposals have in common the assumption that this feature is a form of adaptation to some ecological or physiological condition. On the other hand, the neutralist interpretation states that the variations are due only to different mutational biases. Since all of the traits that have been proposed by the selectionists either appeared to be limited to certain genera or were invalidated by the availability of more data, they cannot be considered as a selective force influencing the genomic GC% across *all* prokaryotes. In this report we show that aerobic prokaryotes display a significant increment in genomic GC% in relation to anaerobic ones. This is the first time that a link between a metabolic character and GC% has been found, independently of phylogenetic relationships and with a statistically significant amount of data.

Key words: Genomic G + C content — Prokaryotes — Mutational bias — Natural selection — Reactive oxygen species

Correspondence to: Héctor Musto; *email*: hmusto@fcien.edu.uy

The genomic GC% of prokaryotes varies from approximately 25% to 75% (Sueoka 1962).There has been a long-standing controversy concerning the causes of this interspecific variation: Is it caused by natural selection or, conversely, is it selectively neutral? The selectionist interpretations have in common the assumption that GC% is a form of adaptation to some ecological or physiological condition. For example, it has been proposed that an increment in GC% could be advantageous for organisms that are exposed to UV radiation (Singer and Ames 1970) and for thermophilic organisms (Argos et al. 1979). Further, it was shown that nitrogen-fixing bacteria display higher GC levels than their non-nitrogen-fixing relatives (McEwan et al. 1998). However, since these proposals appeared to be limited to certain genera or were invalidated by the availability of more data (Galtier and Lobry 1997), they cannot be considered as a selective force influencing GC% across all prokaryotes. Therefore, among the main arguments in favor of the neutralist interpretation is that "no physiological or ecological trait in common among prokaryotes with similar genomic composition has yet been found" (Gautier 2000).

It is well known that aerobic metabolism leads to the formation of reactive oxygen species, which damage different cell components including DNA. The sensitivity of the four bases to reactive oxygen species is different. Thus, it seemed feasible to propose that the relation of the organisms to oxygen could influence its genomic GC%, so that differences

would be found between aerobic and anaerobic species, in particular, among phylogenetically related organisms. Since there are representatives of the two kinds of organisms within several of the known lineages of prokaryotes, we decided to compare their respective genomic compositions. We must note that in 1956 Lee et al. did a similar analysis and could not detect any difference between the two kinds of bacteria, but (a) their data were very limited, (b) facultative bacteria were considered under aerobic organisms, (c) observations of several species of the same genus were considered to be independent.

Figure 1 shows histograms of the genomic GC% of strictly anaerobic (Fig. 1a) and aerobic (Fig. 1b) genera of prokaryotes. A normal distribution of the GC% of anaerobic organisms cannot be rejected (Shapiro–Wilks test, $p < 0.19$), with a mean value of 45%. In contrast, the GC% of aerobic genera was strongly skewed toward high GC% levels, with a mean of 59% and a median of 62.5%. The difference between the two distributions was highly significant. These distributions were not due to an overrepresentation of aerobic (or anaerobic) organisms belonging to the same lineage since both classes of prokaryotes were present within the main branches, as shown schematically in Fig. 2. This result clearly shows that aerobiosis is strongly linked to a significant increment in GC%. Interestingly, the same pattern was evident when aerobic and anaerobic genera were compared within the domains of both Archaea and Bacteria (Table 1). This general trend was even observed in the majority of *phyla* that presented the two types of metabolism (Bacteroidetes, Firmicutes, Proteobacteria, and Euryarcheota) (Table 1). Indeed, the only exception was within Crenarcheota, but we stress that when the difference was significant ($p < 0.05$), the aerobic genera were more GC-rich than the anaerobic ones (Table 1).

These results are unambiguous. However, their interpretation is not trivial, for several reasons. At the DNA level, although reactive oxygen species can attack the four bases and also the sugar residue, it is well established that the most frequently modified base is G. It is the most easily oxidized of the DNA bases according to the reduction potentials of the corresponding radicals, and initial oxidative events occurring in the other bases may be transferred to this base (Steenken and Jovanovic 1997). In addition, G is the base that reacts most rapidly with different oxidizing agents (Ross et al. 1998). The most extensively studied product of G oxidation, and probably the most abundant, is 8-oxo-guanine (Beckman and Ames 1997; Marnett 2000). The importance of this modification is underscored by the fact that many organisms have developed special systems to repair it.

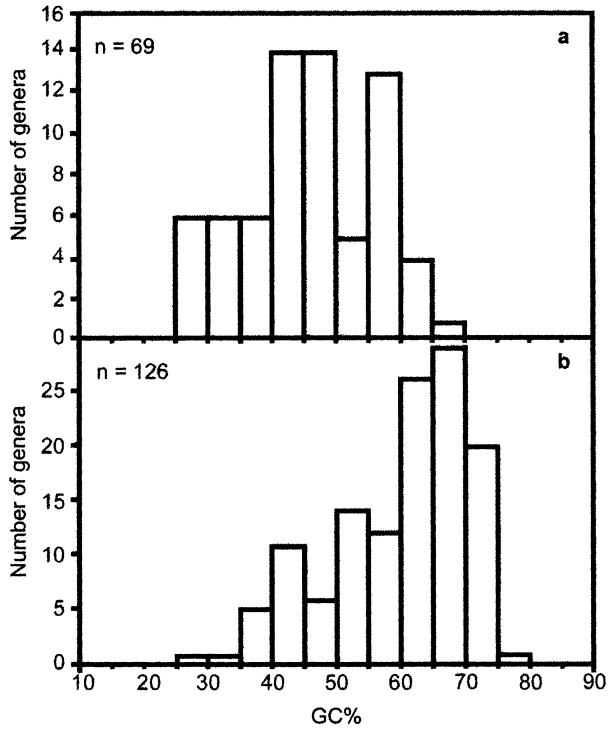

Fig. 1. Histogram of mean GC% from strictly aerobic and anaerobic genera. The organisms and their respective GC% values were taken from Galtier and Lobry (1997) and Holt et al. (1994). For all the species for which the GC% and the relation to oxygen were known, only species defined as strictly aerobic or anaerobic were considered. No genus contained aerobic and anaerobic species. When more than one species from a genus was available, the mean GC% value of the genus was taken. The median value of the standard deviation within a genus was 2.1%, supporting that the mean GC% for a given genus is a valid statistic parameter to use in this analysis. For anaerobic organisms (**a**) the distribution was normal (Shapiro–Wilks test, $p > 0.15$) and centered at 45%, while for aerobic organisms (**b**) the distribution was not normal (mean = 59%, median = 62.5%). The difference between the distributions was significant ($p < 0.0001$, Kruskal–Wallis test). The data are available at http://oeg.fcien.edu.uy/GCprok/.

This modified nucleotide induces G–T transversions, which obviously makes paradoxical the notorious increment in GC% (Cheng et al. 1992; Moriya et al. 1991; Wood et al. 1990).

In line with this point, compilation of mutational spectra obtained by different laboratories demonstrates that GC–AT transitions and GC–TA transversions are the most commonly observed mutations resulting from oxidative damage to DNA (Wang et al. 1998). Across evolution most known mutator genes and repair systems have a strong bias toward AT. Besides, there is a negative correlation between the GC% of mammalian mitochondrial DNAs and the respective metabolic rates (Martin 1995). In other words, as the oxygen consumption is incremented, and hence there are more species that can damage DNA, the AT content of the mitochondria increases. Thus, from a strictly mutationalist point of view, one

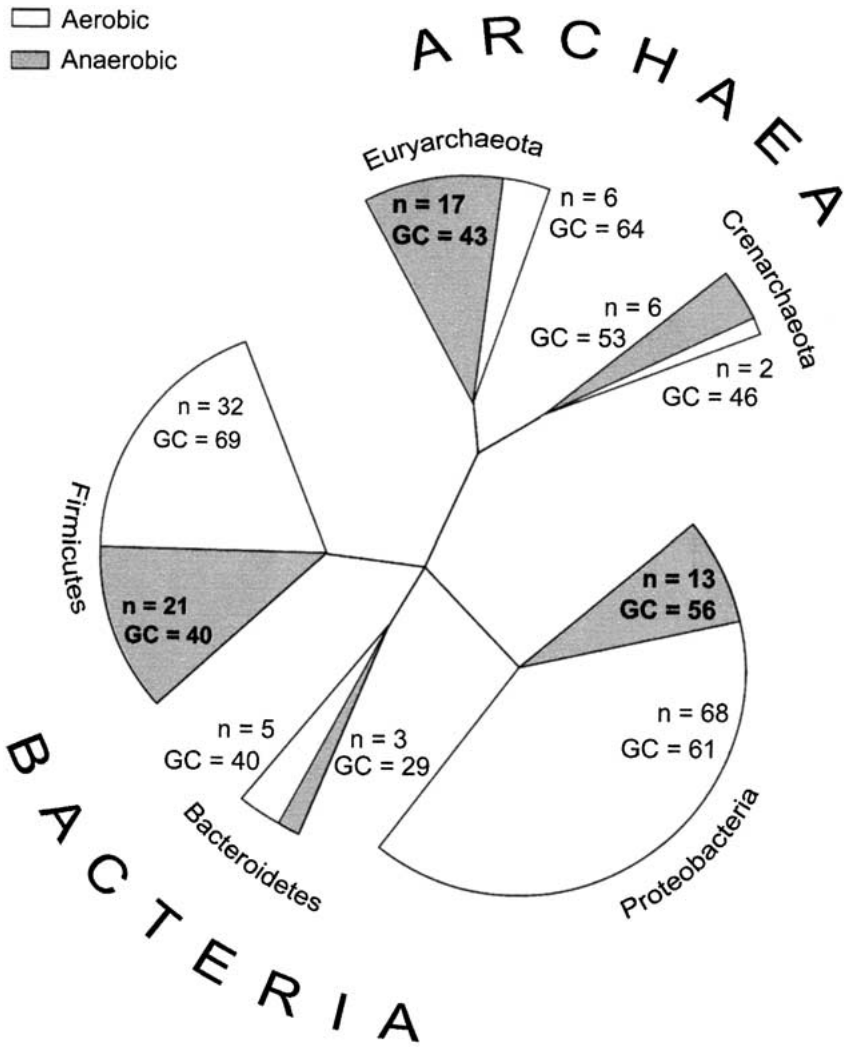

Fig. 2. Diagram of the phylogenetic relationships at the *phylum* level of the aerobic (*white*) and anaerobic (*gray*) prokaryotes included in this study. The mean GC% for each *phylum* is displayed, together with the respective number of genera, *n*.

Table 1. Comparison of GC% in different groups of strictly aerobic and anaerobic genera of prokaryotes[a]

	Aerobic			Anaerobic			
	n	Mean GC%	Median GC%	*n*	Mean GC%	Median GC%	*p*
All	126	59.0	62.0	69	45.5	45.0	< 0.0001
Archaea	8	59.8	64.0	23	45.2	45.0	< 0.002
Euryarcheota	6	64.3	64.0	17	43.1	43.0	< 0.0005
Crenarcheota	2	46.0	46.0	6	51.2	53.5	< 0.6150
Bacteria	118	58.9	62.0	46	45.7	45.0	< 0.0001
Proteobacteria	68	59.0	61.0	13	53.2	56.0	< 0.018
Firmicutes	32	65.1	69.0	21	42.9	40.0	< 0.0001
Bacteroidetes	5	38.6	40.0	3	32.7	29.0	< 0.0460

[a] *n* is the total number of genera considered. *p* is the probability that the two groups (aerobic and anaerobic) have the same median value (Kruskal–Wallis test). The sum of Bacteria (118) does not coincide with its three *phyla* (105) because data for several taxonomic groups with a small number of genera are not displayed.

would expect lower GC levels among aerobic prokaryotes, but as shown, the opposite is true.

One neutralist explanation that could be related to the increment in GC% in aerobiosis is the mutM/mutY system for repair of 8-oxo-guanine, which has been well studied in *Escherichia coli*. This system repairs with a high efficiency when the oxidation occurs in the DNA molecule, but some mutations may result when the oxidation takes place at the level of the dNTP pool (dGTP–8-oxo-dGTP). If the 8-oxo-dGTP

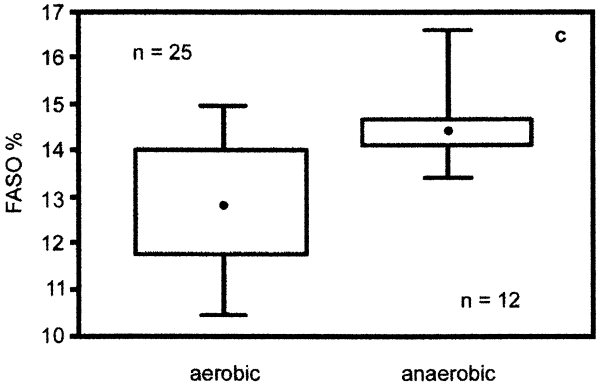

Fig. 3. Box plot of the summed frequencies of amino acids most susceptible to oxidation (FASO), Cys, Met, Trp, Tyr, Phe, and His, for aerobic and anaerobic organisms. The use of amino acids was compared in prokaryotes with more than 100 (aerobic) and more than 50 (anaerobic) nonredundant and complete protein sequences (the difference in the threshold values is due to the scarcity of anaerobic prokaryotes with enough available genes). The two samples had the same genomic GC% distribution as in Fig. 1. Top horizontal bar, maximum; bottom horizontal bar, minimum; box, 25–75%, ●, median. *n*, number of genera.

is incorporated during replication pairing an A, the A will be wrongly removed by MutY, inducing an A–C mutation. Although MutT, a hydrolase, converts 8-oxo-dGTP to 8-oxo-dGMP, preventing the incorporation of the modified nucleotide into the DNA, some of the damaged nucleotides may remain, fixing the mutation and inducing a weak but constant bias toward G:C pairs (Miller 1996). However, some points should be addressed considering the plausibility of this interpretation. First, it assumes that an analogous system with the same bias operates among all prokaryotes. Second, and more important, since the vast majority of mutations is biased toward AT (Wang et al. 1998), the small effect toward incrementing GC should be rapidly counteracted by all other mutations, including the G–T transversions caused by 8-oxo-G (Cheng et al. 1992; Moriya et al. 1991; Wood et al. 1990). Therefore, given the apparent insufficiency of the neutralist interpretation, it is tempting to speculate about selective forces "pushing up" the genomic GC% of aerobic prokaryotes.

We propose that three main factors might be, at least partially, shaping the increment in genomic GC%. First, the pattern of amino acid utilization may be different in aerobic organisms. To investigate this, we analyzed the use of amino acids in all aerobic and anaerobic prokaryotes with at least 100 and 50 available nonredundant protein sequences, respectively. The two samples thus obtained were representative of the two large data sets since each showed the same GC% distribution as in Fig. 1. Remarkably, the summed frequency of the residues Cys, Met, Trp, Tyr, Phe, and His was lower in the genes of aerobic prokaryotes than in those of anaerobic ones ($p < 0.0007$) (Fig. 3). These six amino acids have in common the fact that they are the ones that are oxidized preferentially (Berlett and Stadtman 1997). In turn, the triplets encoding these residues are biased toward AT at the nonsynonymous sites (9 of 14 bases). Particularly, Tyr and Phe (encoded exclusively by A or T at the nonsynonymous sites) are the amino acids with the highest contribution to the differences mentioned above. Thus, an increment in genomic GC%, which involves a decrease in the frequencies of the residues more susceptible to oxidation, could constitute a selective advantage for aerobic prokaryotes.

Second, another selective force could be the fact that high GC% levels increase the thermodynamic stability of the DNA through the formation of three, instead of two, interstrand hydrogen bonds. Less melting of the double helix would in turn diminish the accessibility of oxygen radicals to the bases (Breen and Murphy 1995). In addition, since reactive oxygen species also produce single-strand nicks in the sugar–phosphate backbone of DNA, the increase in GC% could render more stable the DNA structure in the region around a nick.

Third, GC% could be related to the amino acid changes following base substitutions. Indeed GC% is correlated with the frequency of fourfold degenerate codons, which increases from 30% of all triplets in AT-rich genomes to 60% in GC-rich genomes, with a concomitant decrease in twofold degenerate triplets (D' Onofrio et al. 1999). The rise in GC% increases the number of sites at third codon positions in which any mutation is synonymous (fourfold versus twofold degenerate sites). In turn, this decrease in the potential amino acid changes as a result of mutations could offer a selective advantage for aerobic organisms, which are exposed to higher formation rates of DNA-damaging agents. In addition, since G is the most oxidized base, its relative abundance, especially at synonymous sites, could play a sacrificial role, scavenging oxidizing equivalents and protecting the other bases. Given the increment in genomic GC%, most synonymous sites will be occupied by a G, either in the template or in the coding strand of the DNA.

In summary, we have shown that one physiological trait, namely, oxygen metabolism, is linked to the genomic composition of prokaryotes, leading to a remarkable increment in GC% among aerobic species. Since this increment is contrary to the expected mutation pattern, it follows that a high genomic GC% is selectively favored in aerobiosis. Indeed, natural selection could be responsible for the skewed distribution of GC% among aerobic organisms, while the absence of the selective factor (O_2) could account for the normal distribution centered at a nonbiased GC% among anaerobic organisms. We have proposed several selectionist hypotheses to explain these results. Nevertheless, the most relevant issue is our

finding of the link between aerobiosis and GC%, which challenges the accepted neutralist interpretation of the evolution of genomic GC% among prokaryotes.

Acknowledgments. We thank F. Alvarez, A. Castro, B. Garat, K. Jabbari, and J. Oliver for helpful discussions. H.N. was the recipient of a fellowship from PEDECIBA, Uruguay. B.A. was supported by grants from TWAS and CSIC.

References

Argos P, Rossman MG, Grau UM, Zuber H, Frank G, Tratschin JD (1979) Thermal stability and protein structure. Biochemistry 18:5698–5703

Beckman KB, Ames BN (1997) Oxidative decay of DNA. J Biol Chem 272:19633–19636

Berlett BS, Stadtman ER (1997) Protein oxidation in aging, disease, and oxidative stress. J Biol Chem 272:20313–20316

Breen AP, Murphy JA (1995) Reactions of oxyl radicals with DNA. Free Radic Biol Med 18:1033–1077

Cheng KC, Cahill DS, Kasai H, Nishimura S, Loeb LA (1992) 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G—T and A—C substitutions. J Biol Chem 267:166–172

D' Onofrio G, Jabbari K, Musto H, Bernardi G (1999) The correlation of protein hydropathy with the base composition of coding sequences. Gene 238:3–l4

Galtier N, Lobry JR (1997) Relationships between genomic G + C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J Mol Evol 44:632–636

Gautier C (2000) Compositional bias in DNA. Curr Opin Genet Dev 10:656–661

Holt J, Krieg N, Sneath P, Staley J, Stanley (1994) Bergey's manual of determinative bacteriology. William and Wilkins, Baltimore

Lee KY, Wahl R, Barbu E (1956) Contenu en bases puriques et pyrimidiques des acides désoxyribonucléiques des bactéries. Ann Inst Pasteur 91:212–224

Marnett LJ (2000) Oxyradicals and DNA damage. Carcinogenesis 21:361–370

Martin AP (1995) Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. Mol Biol Evol 12:1124–1131

McEwan CE, Gatherer D, McEwan N (1998) Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus. Hereditas 128:179–178

Miller JH (1996) Spontaneous mutators in bacteria: Insights into pathways of mutagenesis and repair. Annu Rev Microbiol 50:625–643

Moriya M, Ou C, Bodepudi V, Johnson F, Takeshita M, Grollman AP (1991) Site-specific mutagenesis using a gapped duplex vector: A study of translesion synthesis past 8-oxodeoxyguanosine in *E. coli*. Mutat Res 254:281–288

Ross AB, Mallard WG, Helman WP, Buxton GV, Huie RT, Neta P (1998) NDRL–NIST Solution Kinetics Database. NDRL–NIST, Notre Dame, IN, Gaithersburg, MD

Singer CE, Ames BN (1970) Sunlight ultraviolet and bacterial DNA base ratios. Science 170:822–825

Steenken S, Jovanovic SV (1997) How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. J Am Chem Soc 119:617–618

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

Wang D, Kreutzer DA, Essigmann JM (1998) Mutagenicity and repair of oxidative DNA damage: Insights from studies using defined lesions. Mutat Res 400:99–115

Wood ML, Dizdaroglu M, Gajewski E, Essigmann JM (1990) Mechanistic studies of ionizing radiation and oxidative mutagenesis: Genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome. Biochemistry 29:7024–7032