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 Musto1

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

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 aerobicism 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 Euryarchaeota) (Table 1). Indeed, the only exception was within Crenarchaeota, 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.

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 transitions 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

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