Hereditas 128: 173–178 (1998)

Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus

CATRIONA E. A. MCEWAN[1,a], DEREK GATHERER[2] and NEIL R. MCEWAN[1,3,b]

[1] *Department of Agricultural Research for Northern Sweden, Crop Science Section, Swedish University of Agricultural Sciences, SE-903 04 Umeå, Sweden*
[2] *School of Biomolecular Sciences, John Moores University, Byrom Street, Liverpool, L3 3AF, England*
[3] *Land Resources Department, Scottish Agricultural College, Craibstone Estate, Bucksburn, Aberdeen, AB21 9SJ, Scotland*

McEwan, C. E. A., Gatherer, D. and McEwan, N. R. 1998. Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus. — *Hereditas 128*: 173–178. Lund, Sweden. ISSN 0018-0661. Received July 7, 1997. Accepted March 4, 1998

The genomic GC contents of both nitrogen-fixing and non-fixing members of eight genera of bacteria are investigated. Analysis by *t*-tests showed that in the two aerobic genera investigated (*Aquaspirillum* and *Vibrio*) there is a significantly higher GC content in the nitrogen-fixing members of the genus than in those unable to fix nitrogen, whilst in anaerobic genera there is either no GC bias, or in the case of two genera (*Rhodospirillum* and *Clostridium*) there is a significantly higher GC content in the non-fixing organisms. This suggests that, in many genera, directional mutational pressures are different in nitrogen-fixing and non-fixing organisms. These results are discussed in the light of known mechanisms of mutation pressure and their relation to environmental variables.

N. R. McEwan, Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, Scotland

Diazotrophy, the ability to fix nitrogen from atmospheric N_2, is found in a wide range of bacterial genera. In some cases, the sole member of the genus is nitrogen fixing (e.g., *Frankia*), in others all members of the genus are nitrogen fixing (e.g., *Beijerinckia*, *Xanthobacter*), but in many cases there is only a single nitrogen-fixing member of the genus (e.g., *Beggiatoa*, *Halobacterium*) with all other members of the genus being unable to fix atmospheric nitrogen. In total, eight different genera have a sufficiently large number of both nitrogen fixing and non-fixing species to permit statistical analysis on any differences between the two groups within the genus. These genera are listed in the materials section, and the species investigated within each genus are listed in Table 1.

Although it is well documented that there is considerable range of genomic GC content across organisms (ranging from about 25 % to well over 70 %; reviewed by LI and GRAUER 1991), the reasons for this range of values has received relatively little attention. To date, four hypotheses have been put forward as potential reasons for creating variation in the GC content.

Firstly, there is the environmental influence of UV irradiation (SINGER and AMES 1970). Since ultraviolet radiation induces the formation of thymine dimers, higher GC contents give a selective advantage to organisms living in niches which are susceptible to direct and intense sunlight.

Secondly, KAGAWA et al. (1984) hypothesise that thermophilic organisms demonstrate a tendency to high GC content since thermostable and thermolabile amino acids are encoded by GC-rich and GC-poor codons respectively. In addition, GC-rich DNA is generally more likely to remain double-stranded at higher temperatures.

Thirdly, deviation from a 50 % GC may be attributed to a deviation in the direction of mutation of AT pairs to GC pairs, and vice versa in a genome (SUEOKA 1988). Unless this ratio is equal to 1, the genomic GC content will drift away from 50 %.

Finally, the differing GC content of closely related species has been considered as a means of discrimination, to permit identification of foreign DNA which might be introduced into the cells by passive transformation (FORSDYKE 1996). In this theory, the selective advantage gained through the easy identification of foreign DNA results in selection pressure for GC content difference in closely related species.

Here we consider a fifth hypothesis, which deals with the implications of diazotrophy on the GC content of the DNA. The number of nitrogen atoms found in each of the five nucleotides is 5, 3, 5, 2, and 2 for adenine, cytosine, guanine, thymine and uracil, respectively. If this is considered in terms of nucleotides which are found in DNA, or which are found in transcribed RNA molecules, rather than free existing nucleotides (e.g., ATP, GMP, etc.), every GC pair has 8 nitrogen molecules to the 7 found in either an AT or AU pairing.

[a] Current Address: Craibstone Estate, Bucksburn, Aberdeen, Scotland
[b] Current Address: Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, Scotland

It therefore might be expected that, where an organism has the capacity to fix nitrogen from the atmosphere, it will be more likely to selectively use a higher GC content than its non-nitrogen fixing relative.

MATERIALS AND METHODS

Only genera where there is a large enough number of both nitrogen fixing and non-fixing species to permit statistical analysis, were used in this study. In total, only eight genera matched this criterion; *Aquaspirillum*, *Chromatium*, *Clostridium*, *Desulfotomaculum*, *Desulfovibrio*, *Methanobacterium*, *Rhodospirillum*, and *Vibrio*. The species studied within these genera are listed in Table 1.

Genomic GC contents were obtained from either BALOWS et al. (1991), BUCHANAN and GIBBONS (1974), or LASKIN and LECHEVALIER (1981).

The distribution of the frequency of GC contents within each genus was plotted for both nitrogen fixing and non-fixing species, using Microsoft Excel. The nitrogen-fixing members of the genus were compared with their non-fixing counterparts by performing *t*-test analysis using Microsoft Excel, with the null hypothesis being that there is no difference in average GC content between the two groups within a genus. The *t*-test selected is two-tailed and either heteroscedastic or homoscedastic depending on whether the variances are unequal or equal, respectively. The *F*-test was applied to detect inequality in variances between groups.

RESULTS

The distribution of genomic GC content is shown in Fig. 1. The average GC content for nitrogen fixers is 49.8 ($n = 39$), and for non-fixers is 41.3 ($n = 159$). Both groups have equal variance as assessed by the *F*-test. A two-tailed equal variance *t*-test confirms that nitrogen-fixing species have a significantly higher GC content than the non-fixing species ($p < 0.0005$). This is in keeping with a hypothesis that nitrogen-fixing organisms will tend to have a higher content of the more N-rich nucleotides (i.e., higher GC content). However, simply to divide the species into fixing and non-fixing categories may obscure any genus specific trends. In particular, any genus specific trends are likely to be masked by there being a distinctly higher number of species found in the *Clostridium* genus, a genus which has species generally AT rich, and will thus elevate the abundance of the lower GC values.

Therefore, species were divided into genera and re-analysed. For individual genera, *F*-tests established significant differences in levels of variance between nitrogen fixers and non-fixers in *Aquaspirillum* ($p < 0.05$) and *Clostridium* ($p < 0.0005$). In *Vibrio*, there is no variance in the non-fixers, since all three species have 46 % GC. Thus for these three genera a heteroscedastic *t*-test was performed.

Initial inspection of the data in Table 2 suggests that in both *Rhodospirillum* and *Clostridium*, the non-fixing organisms tend to have a higher GC content, contrary to the 'global' findings for all organisms, but in *Aquaspirillum* the reverse appears to be true. On performing *t*-test analysis the significance of these observations is confirmed (*Rhodospirillum* and *Clostridium* $p < 0.005$ and *Aquaspirillum* $p < 0.02$). In addition, *Vibrio* is found to have a higher GC content in nitrogen-fixing species, with a confidence value of $p = 0.05$. However, *t*-tests performed on the other genera suggest that the nitrogen fixing and non-fixing species are from common pools in terms of GC content.

DISCUSSION

SUEOKA (1988) defined the concept of mutational pressure, by pointing out that unless the ratio of nucleotide substitutions, for AT pairs mutating to CG pairs and vice versa, in a genome is equal to 1, the GC content will tend to drift away from 50 %. Extreme examples include *Mycoplasma capricolum*, with a GC content of 25 %, and *Micrococcus luteus* at 75 % (reviewed by SUEOKA 1988; LI and GRAUER 1991). The hypothesis presented here is by contrast selectionist in that availability of compounded nitrogen is postulated to influence GC content. Only three other selective hypotheses for GC content have been proposed to date (KAGAWA et al. 1984; SINGER and AMES 1970; and FORSDYKE 1996).

The results presented here demonstrate that there is a highly significant tendency towards higher genomic GC content for those members of *Aquaspirillum*, and to a lesser extent *Vibrio*, which fix nitrogen. However, it is particularly interesting to note that a stronger significance is found in *Rhodospirillum* and *Clostridium*, and these genera both display the opposite bias with nitrogen-fixing species having lower genomic GC contents. *Aquaspirillum* and *Vibrio* are aerobic, whereas the other six genera studied are all anaerobic bacteria. Thus the advantages discussed in the introduction, of having a higher GC content in nitrogen fixing organisms, relative to non-fixing species from the same genus, seems to apply only to aerobic bacteria.

It is unlikely that the selection of GC or AT rich sequences is a reflection of enzymic pathways. For example, both AMP and GMP are derived from a common precursor molecule — IMP. The pathways

Table 1. *The species used in this study*

Aquaspirillum:	N fixing:	*A fasciculus, A. itersonii, A. magnetotacticum, A. peregrinum*
	non-N fixing:	*A. anulus, A. aquaticum, A. autotropicum, A. bengal, A. delicatum, A. dispar, A. giesbergeri, A. gracile, A. metamorphum, A. polymorphum, A. psychrophilum, A. putridiconchylium, A. serpens, A. sinosum*
Chromatium:	N fixing:	*C. gracile, C. minus, C. minutissimun, C. vinosum, C. violascens, C. warmingii, C. weissei*
	non-N fixing:	*C. buderi, C. okenii, C. purpuratum, C. salexigens, C. tepidum*
Clostridium:	N fixing:	*C. aceticum, C. acetobutylicum, C. arcticum, C. beijerinckii, C. butyricum, C. cellobioparum, C. formicoaceticum, C. kluyveri, C. papyrosolvens, C. pasteurianum, C. thermosaccarolyticum*
	non-N fixing:	*C. acidiurici, C. aerotolerans, C. aminovalericum, C. argentinense, C. aurantibutyricum, C. baratii, C. barkeri, C. bifermentans, C. botulinum, C. bryantii, C. butyricum, C. cadaveris, C. carnis, C. celerecrescens, C. cellulolyticum, C. cellulovorans, C. chartatabidum, C. chauvoei, C. clostridiiforme, C. coccoides, C. cochlearium, C. cocleatum, C. colinum, C. collagenovorans, C. cylindrosporum, C. difficile, C. disporicum, C. durum, C. fallax, C. felsineum, C. fervidum, C. ghoni, C. glycolicum, C. haemolyticum, C. indolis, C. innocuum, C. intestinale, C. josui, C. lentocellum, C. leptum, C. limosum, C. lituseburense, C. madisonii, C. magnum, C. malenomanatum, C. methylpentosum, C. nexile, C. novyi, C. oceanicum, C. oroticum, C. oxalicum, C. paraputrificum, C. perfringens, C. pfennigii, C. polysaccharolyticum, C. populeti, C. propionicum, C. proteolyticum, C. puniceum, C. purineinolyticum, C. putrefaciens, C. putrificum, C. quercicolum, C. ramosum, C. rectum, C. saccharolyticum, C. sartagoforme, C. scatolgenes, C. scindens, C. septicum, C. sordellii, C. sphenoides, C. spiroforme, C. sporogenes, C. sporosphaeroides, C. stercorarium, C. sticklandii, C. subterminale, C. symbiosum, C. tertium, C. tetani, C. tetanomorphum, C. thermoaceticum, C. thermoautotrophicum, C. thermobutyricum, C. thermocellum, C. thermocopriae, C. thermohydrosulfuricum, C. thermolacticum, C. thermosulfurogenes, C. tyrobutyricum*
Desulfotomaculum:	N fixing:	*D. nigrificans, D. orientisi, D. ruminis*
	non-N fixing:	*D. acetoxidans, D. antarcticum, D. geothermicum, D. guttoideum, D. kuznetsovii, D. sapomandens, D. thermoacetoxidans*
Desulfovibrio:	N fixing:	*D. africanus, D. desulfuricans, D. gigas, D. salexigans, D. vulgaris*
	non-N fixing:	*D. carbinplicus, D. fructosovorans, D. furfuralis, D. giganteus, D. piger, D. simplex, D. sulfodismutans*
Methanobacterium:	N fixing:	*M. formicium, M. ivanovii, M. thermautrophicum*
	non-N fixing:	*M. alcaliphium, M. bryantii, M. espanolae, M. palustre, M. thermoaggregans, M. thermoalcaliphilum, M. thermoformicicium, M. uliginosum, M. wolfei*
Rhodospirillum:	N fixing:	*R. fulvum, R. molischianum, R. photometricum, R. rubrum, R. salexigens*
	non-N fixing:	*R. centenum, R. mediosalinum, R. salinarum*
Vibrio:	N fixing:	*V. diazotrophicus, V. natriegens, V. pelagius*
	non-N fixing:	*V. aestuarianus, V. albensis, V. alginolyticus, V. anguillarum, V. campbellii, V. cholerea, V. costicola, V. fischeri, V. gazogenes, V. harveyi, V. logei, V. marinus, V. mediterranei, V. metschnikovii, V. nereis, V. nigripulchritudo, V. ordalii, V. orientalis, V. parahaemolyticus, V. proteolyticus, V. psychroerythrus, V. salmonicidia, V. splendidus, V. tubiashii*

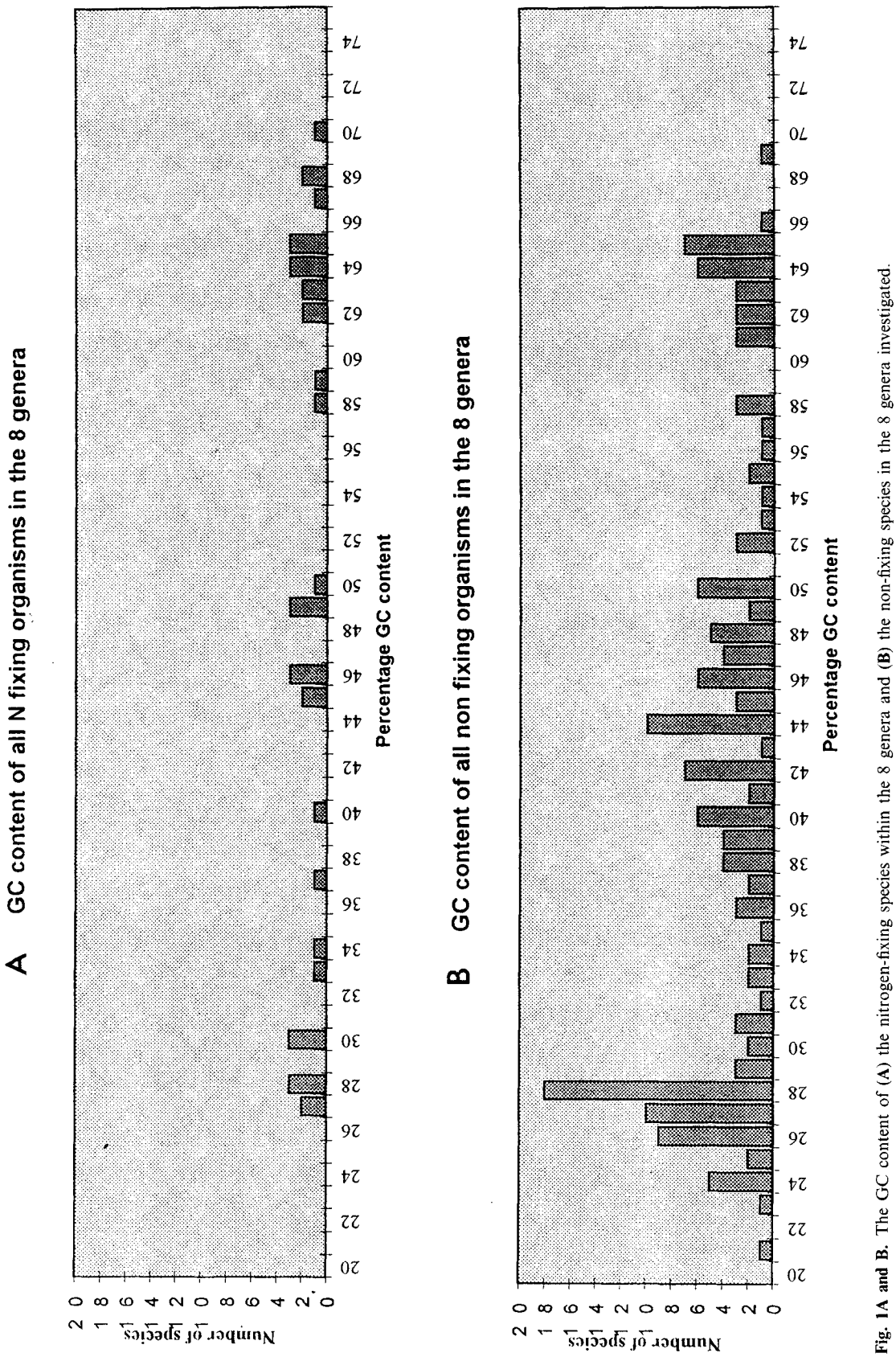

Fig. 1A and B. The GC content of (**A**) the nitrogen-fixing species within the 8 genera and (**B**) the non-fixing species in the 8 genera investigated.

Table 2. *The GC content of the species within each genus separated into nitrogen fixing (N+) and non-fixing (N−) groups: Meth. = Methanobacterium; Chrom. = Chromatium; Desulfov. = Desulfovibrio; Desulfoto. = Desulfotomaculum; Aquasp. = Aquaspirillum; Vib. = Vibrio; Rhod. = Rhodospirillum; Clost. = Clostridium*

GC	Meth.		Chrom.		Desulfov.		Desulfoto.		Aquasp.		Vib.		Rhod.		Clost.	
%	N+	N−	N+	N−	N+	N−	N+	N−	N+	N−	N+	N−	N+	N−	N+	N−
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	10
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	18
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
32	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
34	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
36	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
37	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
38	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3
39	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3
40	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	4
41	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
42	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	3
43	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	4
45	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	2
46	0	0	0	0	0	0	0	0	0	0	3	4	0	0	0	2
47	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
48	0	0	0	0	0	1	0	2	0	0	0	1	0	0	0	1
49	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0
50	1	0	0	0	0	0	0	2	0	1	0	2	0	0	0	1
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1
53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
55	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
56	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
57	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0
59	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0
62	0	0	1	2	0	0	0	0	1	1	0	0	0	0	0	0
63	0	0	1	0	0	0	0	0	1	2	0	0	0	1	0	0
64	0	0	2	0	0	2	0	0	1	3	0	0	0	1	0	0
65	0	0	0	1	2	2	0	0	1	1	0	0	0	3	0	0
66	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
68	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
69	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
70	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
73	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

from IMP to AMP require energy released by the breakdown of GTP, whilst the pathways from IMP to GMP require energy released by the breakdown of ATP to ADP. Both steps have the net effect of liberating a single phosphate bond and require equal amounts of energy. Thus in terms of energy input, both GC and AT pairs have similar contributions from their purine nucleotides.

In a similar way, dTMP may be derived from dCTP, via dUTP, without any change in the energy levels. Thus the net energy input on both GC and AT pairs from their pyrimidine partners is also equal.

An alternative explanation is that nitrogen-fixing anaerobes will only fix nitrogen under conditions when nitrogenous compounds are absent. By definition, anaerobes will not have access to atmospheric oxygen, and it may be possible that there is also a reduction in access to other atmospheric gases. If this were the case, anaerobes which fix nitrogen would be expected to minimise the nitrogen content of their nucleic acids due to a dearth of access to atmospheric nitrogen. On the other hand, their non-fixing relatives will only be able to grow under conditions where there are fixed nitrogen compounds already available; hence they are not subject to the same constraints in terms of GC content. In the case of the aerobic bacteria, the nitrogen-fixing species are assured access to atmospheric nitrogen whilst under conditions of starvation of compounded nitrogen, and so would be relatively more able to use nitrogen-rich nucleotides. Unfortunately, it is not clear that a deficiency of atmospheric oxygen would necessarily lead to a deficiency in atmospheric nitrogen in the niche occupied by the anaerobic genera, and thus this explanation remains speculative.

Nevertheless, what is clear is that there is an elevated GC content in nitrogen-fixing species of aerobic genera, and in the anaerobic genera either the reverse is true, or there is no difference between nitrogen fixing and non-fixing species. Unfortunately, there are insufficient genera where there is a large enough number of species which are able to fix nitrogen and also a large enough number of species which are non-fixers, to permit this study to be expanded.

ACKNOWLEDGEMENTS

The authors wish to thank Dr C. T. Wheeler (University of Glasgow, Scotland) for helpful comments and discussions.

REFERENCES

Balows A, Trüper HG, Dworkin M, Harder W and Schleifer K-H, (1991). The Prokaryotes, 2nd Edition. Springer-Verlag Press.

Buchanan RE and Gibbons NE, (1974). Bergey's Manual of Determinate Bacteriology, 8th Edition. Waverley Press Inc., Baltimore.

Forsdyke DR, (1996). Different biological species "broadcast" their DNAs at different (G + C)% "wavelengths". J. Theor. Biol. 178: 405–417.

Kagawa Y, Nojima H, Nukiwa N, Ishizuka M, Nakajima T, Yasuhara T, Tanaka T and Oshima T, (1984). High guanine plus cytosine content in the third letter of codons of an extreme halophile. J. Biol. Chem. 259: 2956–2960.

Laskin AI and Lechevalier HA, (1981). Handbook of Microbiology, 2nd Edition, Volume III — Microbial Composition: Amino Acids, Proteins, and Nucleic Acids. CRC Press.

Li W-H and Grauer D, (1991). Fundamentals of Molecular Evolution. Sinauer, Sunderland MA.

Singer CE and Ames BN, (1970). Sunlight ultraviolet and bacterial DNA base ratios. Science 170: 822–826.

Sueoka N, (1988). Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA 85: 2653–2657.