

Thermal Adaptation of the Small Subunit Ribosomal RNA Gene: A Comparative Study

Huai-Chun Wang,^{1,2} Xuhua Xia,² Donal Hickey³

¹ Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

² Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

³ Department of Biology, Concordia University, Montreal, Quebec H3G 1M8, Canada

Received: 28 October 2005 / Accepted: 1 March 2006 [Reviewing Editor: Dr. Nicolas Galtier]

Abstract. We carried out a comprehensive survey of small subunit ribosomal RNA sequences from archaean, bacterial, and eukaryotic lineages in order to understand the general patterns of thermal adaptation in the rRNA genes. Within each lineage, we compared sequences from mesophilic, moderately thermophilic, and hyperthermophilic species. We carried out a more detailed study of the archaea, because of the wide range of growth temperatures within this group. Our results confirmed that there is a clear correlation between the GC content of the paired stem regions of the 16S rRNA genes and the optimal growth temperature, and we show that this correlation cannot be explained simply by phylogenetic relatedness among the thermophilic archaeal species. In addition, we found a significant, positive relationship between rRNA stem length and growth temperature. These correlations are found in both bacterial and archaeal rRNA genes. Finally, we compared rRNA sequences from warm-blooded and cold-blooded vertebrates. We found that, while rRNA sequences from the warm-blooded vertebrates have a higher overall GC content than those from the cold-blooded vertebrates, this difference is not concentrated in the paired regions of the molecule, suggesting that thermal adaptation is not the cause of the nucleotide differences between the vertebrate lineages.

Key words: Small subunit ribosomal RNA — Secondary structure — Phylogenetic independent

Correspondence to: Donal Hickey; email: dhickey@alcor.concordia.ca

contrast — GC content — Optimal growth temperature

Introduction

Previous studies have shown that the GC content of 16S ribosomal RNAs is highly correlated with the optimal growth temperature in prokaryotes (Dalggaard and Garrett 1993; Galtier and Lobry 1997; Hurst and Merchant 2001; Wang and Hickey 2002; Nakashima et al. 2003). The increased GC content is concentrated in the stem regions of the molecule, supporting the conclusion that it is the result of natural selection acting to increase the structural stability of the rRNA molecule. In this study, we postulated that structural stability could be increased, not only by a higher GC content in stem regions but also by an increase in the stem length in thermophilic species, because both would increase the number of hydrogen bonds in the stem. Thus we predicted that both GC content and stem length would be increased among the thermophiles.

Another question that is interesting to evolutionary biologists is how quickly the rRNA genes in prokaryotic species can evolve in response to temperature differences. Because of the lack of a universal molecular clock in prokaryotes, we addressed the question by comparing the GC content and length of stems and loops between mesophilic and thermophilic species within the same genus.

We were also concerned about the lack of sample independence which is inherent in cross-species comparisons that simply regress a given phenotypic character on an environmental variable. Since each species is part of a hierarchically structured phylogeny, they cannot be regarded as independent samples from the same distribution (Felsenstein 1985; Harvey and Pagel 1991). Therefore, in this study we controlled for phylogenetic relatedness among the archaeal species by using phylogeny-based comparative methods.

Our main focus was on archaeal species, because this group is characterized by the greatest diversity of optimal growth temperatures. In addition, however, we did separate analyses of both bacterial and vertebrate small subunit rRNAs. This allowed us to assess whether the patterns of nucleotide content and stem length that we observed when comparing mesophilic and thermophilic archaea were repeated in the other kingdoms of life.

Materials and Methods

Sequence Data

The 16S rRNA sequence data were downloaded from the SSU rRNA database <http://oberon.fvms.ugent.be:8080/rRNA/ssu/index.html> (Wuyts et al. 2002). For prokaryotes, we retrieved 10,566 bacterial and 590 archaeal 16S rRNA sequences. For each rRNA sequence the GC content and length were calculated. In addition to calculating the total GC content and sequence length for each rRNA, we also calculated these values separately for the stem and loop regions, as defined structurally in the SSU rRNA database.

Since many species have multiple entries in the 16S rRNA sequence database, the data were filtered by retaining only the first entry for each species, which left rRNA data for a total of 4598 species. (When all sequences for a single species were retained and the mean values for each species were used, the subsequent analyses gave very similar results as the above method to filter multiple entries of same species.) 18S rRNA sequences of 84 vertebrate species were also collected, of which 38 species are warm-blooded animals including 34 birds and 4 mammals, and 36 species are cold-blooded animals including 23 fish, 18 amphibians, and 5 reptiles.

Growth Temperature

Optimal growth temperatures of 9168 bacterial and archaeobacterial species were kindly provided by Dr. Manfred Kracht of German Collection of Microorganisms and Cell Cultures (<http://www.dsmz.de>). Many species are listed more than once in the database, sometimes with different growth temperatures. For example, there are 101 temperature entries for *Bacillus* sp., 82 of which have a temperature of 30°C. In other cases the temperatures are all the same or similar. For simplicity the temperature of first entry for each species was assigned for that species, which leaves temperature data for 3630 unique species. We also calculated average temperature for the same species if there were multiple entries. In the case of *Bacillus* sp. for instance, the average temperature is 34.49°C. The correlation between the two temperature data sets is 0.98 ($p < 0.00001$). There is no marked difference in subsequent analyses when using the two temperature data sets. The temperature data were then merged with the rRNA data, and 1673 prokaryotic species that have both temperature and rRNA data

(GC content and gene length) were obtained. Of these, 1573 species are bacterial and the other 100 are archaeal.

Statistical Analyses

We used the nonparametric Wilcoxon rank sum tests to compare average GC content and length between two temperature groups such as mesophiles and thermophiles in bacteria. The Kruskal-Wallis rank sum tests were used to compare the data among three temperature groups: mesophilic, moderately thermophilic, and hyperthermophilic bacteria.

For the archaeal species we obtained phylogenetically independent contrasts for temperature and GC content and, also, for temperature and sequence length. These contrasts were calculated using the CONTRAST program (Felsenstein 1985) implemented in the PHYLIP program package, version 3.6 (beta release; <http://evolution.genetics.washington.edu/phylip.html>). The phylogenetic relationships are based on two recently published archaeal trees (Brochier et al. 2004). One of these trees was constructed from the concatenated sequences of transcription proteins (such as RNA polymerases) of 20 archaeal species whose complete or near-complete genomes have been sequenced. The other tree was derived from a concatenation of ribosomal proteins of 18 archaeal species for which complete genome sequences are available. The reason that protein trees rather than an rRNA tree were used here is twofold. First, it is known that phylogeny based on rRNA can be biased when the GC content of the rRNAs is very different (e.g., Hasegawa and Hashimoto 1993; Foster and Hickey 1999). Second, since we are studying the composition of rRNA sequences, it is necessary to use a phylogeny that is not based on these sequences themselves.

16S rRNA sequences for the 20 archaeal species were extracted from the SSU rRNA database (Wuyts et al. 2002) or GenBank if they were not available in the rRNA database. For *Methanosarcina acetivorans*, only a partial sequence of 16S rRNA (GenBank locus MESRR16SA; 1426 bp) was retrieved from GenBank. Three copies of 16S rRNA gene were annotated (MA0896, MA1618, MA4614) in the complete genome sequence *M. acetivorans* (Galagan JE et al. 2002) and they are also partial (1427 bp each).

Results

This study consisted of a number of complementary analyses of 16S and 18S rRNA sequences. First, we did a broad survey of the average trends in mesophilic, thermophilic, and hyperthermophilic species of prokaryotes. Second, we performed a more detailed analysis of the archaeal sequences, with particular emphasis on the potential confounding effects of phylogenetic relatedness. Third, we looked for adaptive divergence between sequences from closely related species of bacteria, in an attempt to estimate the evolutionary time that is required to achieve adaptive change in these sequences. Finally, we examined rRNA sequences from warm-blooded and cold-blooded vertebrate species in order to see if the same adaptive patterns could be seen in these lineages.

Parallel Trends in the Nucleotide Composition of Archaea and Bacteria

For this analysis, species were grouped according to their optimal growth temperatures. We defined the

Table 1. The average GC content and sequence length, \pm standard error, of 16S rRNA stems and loops for mesophilic ($<40^{\circ}\text{C}$), moderately thermophilic ($40^{\circ}\text{--}75^{\circ}\text{C}$), and hyperthermophilic ($\geq 75^{\circ}\text{C}$) bacteria and archaea

	Stem regions			Loop regions		
	Mesophiles	Thermophiles	Hyperthermophiles	Mesophiles	Thermophiles	Hyperthermophiles
(A) Nucleotide composition						
Bacteria ($n = 1573$)	66.1 ± 0.11	70.1 ± 0.37	76.5 ± 1.55	40.1 ± 0.04	41.6 ± 0.17	41.2 ± 0.35
Archaea ($n = 100$)	67.4 ± 0.26	70.8 ± 1.62	81.0 ± 0.41	38.97 ± 0.21	39.2 ± 0.57	42.1 ± 0.25
(B) Sequence length						
Bacteria ($n = 1573$)	885.2 ± 0.52	894.9 ± 2.88	922.2 ± 6.87	637.3 ± 0.39	640.6 ± 2.44	634.3 ± 4.14
Archaea ($n = 100$)	881.3 ± 1.83	891.1 ± 4.55	903.2 ± 3.03	590.2 ± 1.17	590.9 ± 2.38	589.4 ± 2.01

following three groups: the mesophile group contained species with optimal growth temperatures less than 40°C , in which 1461 species are bacterial and 61 species are archaeal; the thermophile group included those species with growth temperatures between 40° and 75°C , in which 106 species are bacterial and 8 species are archaeal; and the hyperthermophile group contained all species with a growth temperature higher than 75°C , with 6 species being bacterial and 31 species archaeal. The boundaries of these groupings are arbitrary; they were defined to create three major “bins” that spanned the entire range of growth temperatures and that contained several species in each group. For each group, the archaeal and bacterial species were analyzed separately.

Rather than simply calculating the GC content of the entire rRNA sequences for each group, we partitioned the sequences into the double-stranded stem regions and the single-stranded loop regions. The results are shown in Table 1A. From the table, we can see that the GC content of the stem regions is highest in the hyperthermophiles, intermediate in the moderate thermophiles, and lowest in the mesophiles. This is true of both the archaeal and the bacterial groups. The Kruskal-Wallis rank sum tests indicated that these differences are highly significant ($p < 0.0001$). In contrast to the stem regions, the nucleotide content of the loop regions does not differ greatly among the three groups and there is no obvious correlation between the GC content of the loops and the optimal growth temperature. Overall, these results confirm previous reports that the GC content of stem regions is positively correlated with the optimal growth temperature, while the loop regions of the same molecules show little or no correlation.

Parallel Trends in Sequence Length of Archaea and Bacteria

Table 1B shows the results of a different analysis of the same species groups as reported in Table 1A. In this case, we calculated the average cumulative sequence length of the stem and loop regions of the rRNA, rather than the average GC content as in

Table 1A above. We see that the average stem length is greatest among the hyperthermophiles, intermediate among the moderate thermophiles, and shortest among the mesophiles. Again, this is true for both the archaeal and the bacterial groups, and the differences are highly significant based on the Kruskal-Wallis rank tests. In contrast to the stem regions, the lengths of the loop regions do not show the same trends (see Table 1B) and the differences between groups are not statistically significant.

Taken together, these results confirm our prediction that thermophilic archaea and bacteria are selected for both higher GC content and longer sequence length in the stem regions. Our prediction was based on the fact that both factors (GC content and length) would result in an increased number of hydrogen bonds, thus increasing the stability of the stems at higher growth temperatures.

It is interesting to note that despite the highly parallel trends in the stem regions of archaeal and bacterial rRNAs, there is a significant difference in the overall length of the rRNA molecule between the two groups. Specifically, the bacterial rRNAs are longer than their archaeal homologues. This overall length difference has already been noted (Wuyts et al. 2002) and can be explained by the fact that the bacterial rRNAs have two extra hairpin loop structures. As can be seen from Table 1B, most of the length difference between the archaeal and the bacterial rRNAs is in unpaired loop regions of the molecule.

Phylogeny-Based Analysis of Thermal Adaptation in Archaeal 16S rRNA

While the vast majority of bacterial species are mesophiles with optimal growth temperatures below 40°C , the archaea span a much wider range of growth temperatures. For this reason we performed a more detailed analysis of the archaeal sequences. First, we plotted the GC content of 100 archaeal rRNA sequences against the optimal growth temperatures of the individual species (see Fig. 1). From Fig. 1 we can see the apparent strong correlation between the GC content of the stem regions and the optimal growth

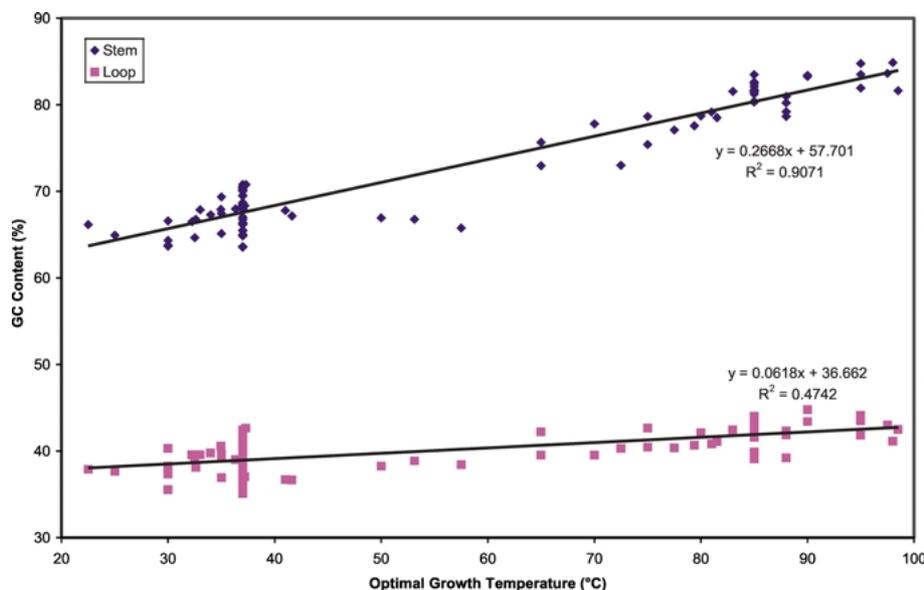


Fig. 1. Relationship between the GC content of 16S rRNA stems and loops and the optimal growth temperature. The data are plotted for 100 archaeal species for which both complete 16S rRNA sequences and optimal growth temperatures are known. The data for the paired stem regions are shown as blue diamonds, while the data for the single-stranded loop regions are shown as purple squares. For a full list of species names and growth temperatures, see Supplementary Table 2.

Table 2. Relationship between (i) GC content and optimal growth temperature and (ii) rRNA sequence length and optimal temperature

	GC content	Sequence length
Original data	0.90*	0.79*
Contrasts based on translation tree	0.85*	0.84*
Contrasts based on transcription tree	0.84*	0.44 ($p = 0.043$)

Note. The values shown are correlation coefficients. Data from 20 species were used for the original dataset and for the contrasts based on the translation proteins; data from 18 species were used for the contrasts based on the transcription proteins (see text for explanation). * $p < 0.001$.

temperature. We were concerned, however, that this apparent correlation might be due to phylogenetic clumping of the species. In order to eliminate this possibility, we derived phylogenetically independent contrasts (see Methods) for (i) GC content of the 16S rRNA sequence and growth temperature and (ii) sequence length and growth temperature. The results are given in Table 2. It is clear that the correlations remain when phylogenetic relationships are taken into account. Thus, we can conclude that the differences in both sequence composition and sequence length are due to natural selection rather than historical accident. One of the correlations for the contrasts based on sequence length and temperature is much lower than the others (see Table 2). This is the result based on the phylogeny that was built using transcription related proteins. Upon further investigation, we found that this lowered correlation was due to two outlier points associated with three *Methanosarcina* species. Furthermore, we found that the database entry for *M. acetivorans* is a partial

rRNA sequence (from GenBank) which has only 1426 bases, while the closely related *M. mazei* is represented by a full-length sequence of 1474 bases. This results in a huge contrast both between these two species and between the ancestral node (*M. acetivorans* and *M. mazei*) and the node for *M. barkeri*. When the two outlier points were removed the correlation between the rRNA length contrasts and the temperature contrasts becomes much stronger ($R = 0.88$, $p < 0.001$). This is consistent with the high correlations found when using the tree based on translation proteins—since this tree did not contain the *Methanosarcina* species.

Can We Detect Thermal Adaptation in Phylogenetically Related rRNA Sequences?

The analyses described above show that ribosomal RNA sequences can change as a result of temperature-based natural selection, but they provide little idea of how quickly such adaptation can occur. To address the latter question about the rate of adaptation, we screened the 1573 bacterial species in the rRNA database for cases where a single genus contained both mesophilic and thermophilic species. Of a total of 444 genera, we found 19 genera each containing at least one mesophilic species and one thermophilic species (optimal temperature, $\geq 45^\circ\text{C}$). Within each genus, we compared the GC contents and sequence lengths of the rRNA stems between the mesophilic and the thermophilic species. The results are given in Table 3. We found that the GC content of the rRNA stems was higher in the thermophilic species than in the mesophilic species ($p < 0.001$ for a paired two-tail t -test). The stem lengths are also longer in the thermophilic species than in the mesophilic species ($p = 0.018$, paired t -test), while the

Table 3. Average optimal growth temperature, GC content (%), and length (bases) of 16S rRNA stems of mesophilic species (Meso) and thermophilic species (Thermo) in a bacterial genus

Genus	Temperature (°C)		GC content		Sequence length	
	Meso	Thermo	Meso	Thermo	Meso	Thermo
<i>Actinomadura</i> (18,1) ^a	30.74	55	73.09	73.56	873	876
<i>Actinopolyspora</i> (1,1)	37	45	73.39	74.94	909	916
<i>Amycolatopsis</i> (7,1)	28.29	45	71.21	71.77	870	883
<i>Bacillus</i> (46,7)	30.03	55	64.6	68.9	907	903
<i>Brevibacillus</i> (6,2)	31.65	47.5	65.49	66.62	903	912
<i>Clostridium</i> (101,4)	34.57	57.34	63.07	63.19	872	899
<i>Deinococcus</i> (1,2)	30	47.5	66.59	70.21	874	878
<i>Desulfotomaculum</i> (6,8)	34.08	56.25	64.08	68.93	895	901
<i>Lactobacillus</i> (45,1)	32.97	45	61.49	62.9	913	911
<i>Mycobacterium</i> (43,2)	35.6	45.5	70.25	72.07	884	885
<i>Porphyrobacter</i> (1,2)	30	45	65.21	65.43	845	852
<i>Pseudonocardia</i> (7,1)	28.64	45	72.22	72.34	874	882
<i>Rubrobacter</i> (1,1)	37	50	69.82	76.15	888	888
<i>Saccharomonospora</i> (4,2)	29.42	47.43	71.85	73.06	865	867
<i>Saccharopolyspora</i> (5,1)	28	51.88	70.89	71.87	873	865
<i>Spirochaeta</i> (7,1)	32.71	65	65.8	75	891	920
<i>Streptomyces</i> (67,5)	28.26	46.5	71.36	72.87	885	891
<i>Sulfobacillus</i> (1,1)	35	45	69.5	74.29	887	884
<i>Thermoactinomyces</i> (1,5)	35	51.08	67.39	70.82	897	898
Mean ± SE	32.05 ± 0.7	49.8 ± 1.3	68.3 ± 0.84	70.8 ± 0.89	884.5 ± 4.0	890.1 ± 4.20

^aThe first number in parentheses is the number of mesophilic species and the second number is the number of thermophilic species in the genus. For a full list of species names and growth temperatures, see Supplementary Table 1.

difference in loop length is not significant ($p = 0.62$, paired t -test) for the 19 genera (data not shown). Despite the statistical significance of the differences in the within-genus comparisons, the magnitude of the differences is much less than that in the broader phylogenetic comparisons (see Table 1 and Fig. 1). This reflects both the shorter evolutionary time involved and also the fact that the growth temperature differences are usually not very large within a single genus. One genus in which there are large differences in optimal growth temperatures is *Spirochaeta*, and in this case we see correspondingly large divergences in GC content and sequence length (see Table 3).

Patterns of Nucleotide Composition in Vertebrate 18S rRNA

Having demonstrated the effect of growth temperature on the nucleotide composition and sequence length for prokaryotic 16S rRNA, we asked if a similar trend can be seen in vertebrates when we compare 18S rRNA sequences from homeotherms and poikilotherms. One might expect the warm-blooded homeotherms (mammals and birds) to show contrasting patterns to the cold-blooded vertebrates (fish, amphibians, and reptiles). Indeed, overall genomic differences in GC content have already been noted for warm-blooded and cold-blooded vertebrates (Bernardi 2000). In agreement with these genomic averages, we found that the average GC content of 18S

rRNA sequences among the warm-blooded mammals and birds, approximately 55.7%, is moderately higher than that of the cold-blooded vertebrates, approximately 53.5% (see Table 4). When we partition the 18S sequences into stem and loop regions, however, we see that, in contrast to the prokaryotic data set, the differences are not concentrated in the stem regions of the molecule. For example, the GC content of the rRNA stems in warm-blooded birds is closer to that of the cold-blooded amphibians than it is to the warm-blooded mammals (Table 4).

Variations in rRNA sequence length are also reported in Table 4. Again, there is no obvious grouping of the warm-blooded and cold-blooded classes. Instead, there is a highly significant difference in overall sequence length within the warm-blooded vertebrates, i.e., between mammals and birds. Specifically, the 18S rRNA length is greater in mammals than in birds and the difference is highly significant ($p < 0.0001$), whereas the average length difference between the warm- and the cold-blooded vertebrates is only marginally significant ($p = 0.049$). When we score the stem and loop lengths separately, we see even more variation within the groups and no overall difference between them. For instance, the cold-blooded fish have a longer stem length than the warm-blooded birds. It is interesting to note that the increase in length of the vertebrate 18S rRNA compared to the prokaryotic 16S rRNA is mainly in the loop regions of the molecules.

Table 4. GC content and sequence length of 18S rRNA among five classes of vertebrates

Vertebrate class	Sequence composition (GC%)			Sequence length (bases)		
	Stems	Loops	Average	Stems	Loops	Total
Fish ($n = 23$)	62.1 ± 0.22^a	43.8 ± 0.28	53.4 ± 0.23	956.3 ± 2.47	869.0 ± 1.78	1825.3 ± 2.83
Amphibian ($n = 18$)	62.6 ± 0.21	42.5 ± 0.31	53.0 ± 0.24	901.1 ± 9.37	927.4 ± 8.04	1828.6 ± 6.18
Reptiles ($n = 5$)	61.8 ± 0.21	44.5 ± 0.21	53.6 ± 0.15	900.4 ± 14.5	913.0 ± 13.06	1813.4 ± 1.66
Aves ($n = 34$)	63.2 ± 0.08	47.4 ± 0.17	55.6 ± 0.12	945.6 ± 3.37	876.9 ± 2.96	1822.4 ± 0.48
Mammal ($n = 4$)	64.1 ± 0.19	46.4 ± 0.12	55.8 ± 0.18	998.6 ± 4.75	870.5 ± 3.38	1869.3 ± 2.32

^aMean \pm standard error.

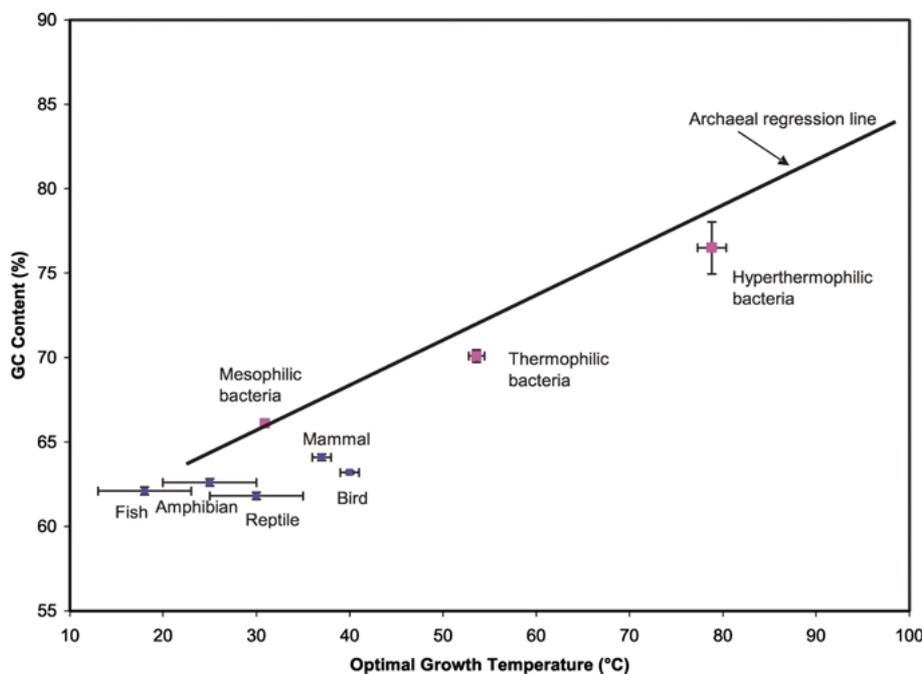


Fig. 2. Relationship between GC content of small subunit rRNA stems and growth temperature. The trend line for the archaeal species (see Fig. 1) is repeated here for reference. The averages for the three groups of bacteria (mesophilic, thermophilic, and hyperthermophilic) are shown as purple squares. The averages for the five vertebrate classes are shown as blue circles. Standard errors are included.

Discussion

In this study, we confirmed the previous findings of a strong positive correlation between the GC content of 16S ribosomal RNA and the environmental growth temperature, and we showed that this correlation is not merely an artifact of phylogenetic relatedness. We also found evidence of a positive relationship between the length of rRNA stems and the temperature. Since the same patterns were repeated within the archaeal and bacterial lineages (see Table 1 and Fig. 2), we can conclude that the positive relationship is not due to phylogenetic history but reflects a repeated selective response to elevated environmental temperature. This conclusion is supported by the phylogenetically independent contrasts of archaeal species. The fact that the increased GC content and increased sequence length are both concentrated in the paired stem regions of the molecule provides further evidence for increasing selection pressure to maintain the folded

structure of the rRNA molecule at higher temperatures.

Taken together, these results indicate that prokaryotic 16S rRNAs respond to increased environmental growth temperature by increasing the structural stability of their rRNAs. This is achieved by increasing both the GC content and the length of the paired regions. Both of these factors (increased GC and increased length of paired regions) increase the number of hydrogen bonds between the paired strands. Thus it is reasonable to interpret these changes as adaptations to growth at high temperature. A previous study (Galtier and Lobry 1997) has shown that the GC contents of several structural RNAs in prokaryotes also positively correlated with the optimal growth temperature. Based on the results presented here for the 16S small subunit rRNA, it is tempting to speculate that the stem lengths of the large subunit 23S rRNA genes are also correlated with the optimal growth temperature. This remains to

be confirmed, however, once a larger database of 23S rRNA sequences becomes available.

In contrast to the data on mesophilic and thermophilic prokaryotes, the differences between warm-blooded and cold-blooded vertebrates are not large, and more significantly, they are not concentrated in the paired stem regions of the molecule (see Table 4). At first glance, this seems to contradict the findings on prokaryotes, but when we consider that the body temperatures of mammals (37°C) and birds (39°C) are not even within the “moderately thermophilic” range of prokaryotes, the result is not at all surprising. Moreover, given that most species of “cold-blooded” snakes prefer to maintain an active body heat at about 30°C—and up to 40°C for desert reptiles—the temperature differences between the two groups are not very clear-cut. In fact, there is a much greater difference between the average body temperature of fish and reptiles than there is between reptiles and mammals. Thus, the results for the vertebrates, although negative, are entirely consistent with the results for the bacteria and archaea.

Despite the lack of obvious temperature-induced differences between the 18S rRNA sequences of warm-blooded and cold-blooded vertebrates, these sequences do illustrate many of the same general features seen in the prokaryotic 16S rRNAs. For instance, the paired regions are relatively GC-rich, while the rRNA loops of all five groups of vertebrates have a very high content of adenine (Wang 2005). This is reminiscent of the elevated amount of adenine in prokaryotic 16S rRNA loops in both mesophilic and thermophilic prokaryotes (Wang and Hickey 2002). In fact, this compositional bias also exists in other eukaryotic species, including protists, fungi, invertebrates, and plants (Wang 2005). There is existing evidence that adenine contributes to the stability of the single-stranded regions of the molecule (Gutell et al. 2000).

Finally, the fact that we see small but significant divergences in sequence composition between related bacteria that have contrasting optimal growth temperatures indicates that these molecular adaptations to the elevated growth temperatures can occur over a relatively short evolutionary time span.

Acknowledgments. This work was supported by a Research Grant from NSERC Canada (D.A.H.) and an Ontario Graduate Scholarship (H.C.W.). We thank Dr. N. Galtier and two reviewers for their comments.

References

- Bernardi G (2000) The compositional evolution of vertebrate genomes. *Gene* 259:31–43
- Brochier C, Forterre P, Gribaldo S (2004) Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol* 5(3):R17
- Dalgaard JZ, Garrett RA (1993) Archaeal hyperthermophile genes. In: Kates M, Kushner DJ, Matheson AT (eds) *The biochemistry of Archaea (Archaeobacteria)*. Elsevier, Amsterdam, p 535
- Felsenstein J (1985) Phylogeny and the comparative method. *Am Nat* 125:1–15
- Foster PG, Hickey DA (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J Mol Evol* 48:284–290
- Galagan JE, Nusbaum C, Roy A, et al. (2002) The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* 12:532–542
- Galtier N, Lobry JR (1997) Relationships between genomic GC content, RNA secondary structures and optimal growth temperature in prokaryotes. *J Mol Evol* 44:632–636
- Gutell RR, Cannone JJ, Shang Z, Du Y, Serra MJ (2000) A story: unpaired adenosine bases in ribosomal RNAs. *J Mol Biol* 304:335–354
- Harvey PH, Pagel MD (1991) *The comparative method in evolutionary biology*. Oxford University Press, Oxford
- Hasegawa M, Hashimoto T (1993) Ribosomal RNA trees misleading? *Nature* 361:23
- Hurst LD, Merchant AR (2001) High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. *Proc R Soc Lond B* 268:493–497
- Nakashima H, Fukuchi S, Nishikawa K (2003) Compositional changes in RNA, DNA and proteins for bacterial adaptation to higher and lower temperatures. *J Biochem* 133:507–513
- Singer GAC, Hickey DA (2003) Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. *Gene* 317:39–47
- Van de Peer Y, De Rijk P, Wuyts J, Winkelmans T, De Wachter R (2000) The European small subunit ribosomal RNA database. *Nucleic Acids Res.* 28:175–176
- Wang H-C (2005) The effects of nucleotide bias on genome evolution. PhD thesis. University of Ottawa, Ottawa
- Wang H-C, Hickey DA (2002) Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Res* 30:2501–2507
- Wuyts J, Van de Peer Y, Winkelmans T, De Wachter R (2002) The European database on small subunit ribosomal RNA. *Nucleic Acids Res* 30:183–185