Significant Correlation between Growth Temperature and Guanine-Cytosine Content in Bacteria

En-Ze Hu†, Xin-Ran Lan†, Zhi-Ling Liu, Jie Gao, and Deng-Ke Niu*

MOE Key Laboratory for Biodiversity Science and Ecological Engineering and Beijing Key Laboratory of Gene Resource and Molecular Development, College of Life Sciences, Beijing Normal University, Beijing, 100875, China

†These authors contributed equally to this paper.

*Corresponding author: Deng-Ke Niu, dkniu@bnu.edu.cn, dengkeniu@hotmail.com

Abstract

Because GC pairs are more stable than AT pairs, GC-rich sequences were proposed to be more adapted to high temperatures than AT-rich sequences. Previous studies consistently showed positive correlations between the growth temperature and the GC contents of structural RNA genes. However, for the whole genome sequences and the silent sites of the codons in protein-coding genes, the relationship between GC content and growth temperature is in a long-lasting debate. With a dataset much larger than previous studies (681 bacteria and 155 archaea), our phylogenetic comparative analyses showed positive correlations between optimal growth temperature and GC content exists, both in the structural RNA genes of bacteria and archaea and in bacterial whole genome sequences, chromosomal sequences, plasmid sequences, core genes, and accessory genes. However, in the 155 archaea, we did not observe a significant positive correlation of optimal growth temperature with whole-genome GC content or GC content at four-fold degenerate sites. We randomly drew 155 samples from the 681 bacteria for 1000 rounds. In most cases (> 95%), the positive correlations between optimal growth temperature and genomic GC content became statistically nonsignificant ($P > 0.05$). This result suggested that the small sample sizes might account for the lack of positive correlations between growth temperature and genomic GC content in the 155 archaea and the bacterial samples of previous studies.

Key words: GC content, optimal growth temperature, prokaryotes, phylogenetic generalized least squares (PGLS), resampling analysis.

Introduction

As guanine (G) strictly pairs with cytosine and adenine (A) pairs with thymine (T) in DNA double helix, the amount of G is equal to C, and that of A is equal to T in the genomes of any cellular organisms. GC content, i.e., the percentage of G+C, is widely used as a measure of genomic

nucleotide composition. It is a highly variable trait ranging from 8% to 75% (Basak et al. 2010; Nguyen et al. 2020). This genomic trait has been widely studied, and its evolution has been proposed to be associated with numerous mutational and selective forces driven by genetic, metabolic, and ecological factors (Foerstner et al. 2005; Hildebrand et al. 2010; Mann and Chen 2010; Raghavan et al. 2012; Wu et al. 2012; Agashe and Shankar 2014; Glemin et al. 2014; Šmarda et al. 2014; Reichenberger et al. 2015; Aslam et al. 2019; Dietel et al. 2019; Weissman et al. 2019). Among these factors, the high temperature might be the most long-debating one (Meyer 2021). Because G:C pairs have an additional hydrogen bond than A:T pairs, the GC-rich genomes are thermally stable and thus more adapted to high-temperature environments. Bernardi and Bernardi (1986) proposed this hypothesis to explain the higher GC content of warm-blooded animals.

As prokaryotes have a much wider thermal distribution on earth than plants and animals, bacterial and archaeal genomes are the best materials to test the thermal adaptation hypothesis. An analysis of 764 prokaryotic species, including mesophilic genera and thermophilic genera, did not find a correlation between whole-genome GC content (GC_w) and the optimal growth temperature (Topt) (Galtier and Lobry 1997). However, this study found a significant positive correlation between Topt and the GC content of structural RNA (tRNAs and rRNAs). The rationale of these observations is that the secondary structures of tRNAs and rRNAs are more sensitive to elevated temperature than the double-strand helix of DNA. In most prokaryotes, protein-coding genes take most of the genome size. Protein structures and functions constrain the GC content at the nonsynonymous sites of the codons. This functional constraint might conceal the hypothetical thermal adaptation. Compared with GC_w, the GC content at the third sites of the codons (GC_3) is more desirable to test the thermal adaptation hypothesis. Early solitary cases indicated that GC_3 might be related to growth temperature. For example, the tyrosy1-tRNA synthetase gene isolated from the thermophile *Bacillus stearothermophilus* has a higher GC_3 than the homologous gene in *Escherichia coli*, 68.0% *vs.* 59.4% (Winter et al. 1983). The *leuB* gene isolated from the extreme thermophile *Thermus thermophilus* HB8 has an extremely high GC_3, 89.4% (Kagawa et al. 1984). For a general conclusion, Hurst and Merchant (2001) examined the relationship between GC_3 and Topt of 29 archaeal species and 72 bacterial species. Unfortunately, they found that Topt is not significantly correlated with GC_3 or GC_w. At the same time, they also found a significant positive correlation between the GC content of structural RNAs and the Topt in both archaea and bacteria. As their analysis had accounted for the effect of shared ancestry, they provided more robust evidence against the thermal adaptation hypothesis. Soon afterward, Xia et al. (2002) showed that growth at increasing temperature (from 37°C to 45°C) for 14,400 generations did not increase but decreased the genomic GC content of the bacterium *Pasteurella multocida*. Meanwhile, the GC contents of two genes, *ldh-a* and *α-actin*, have been analysed across 51 vertebrate species with adaptation temperatures ranging from −1.86°C to approximately 45°C. No significant positive correlation between living temperature and GC content was found, no matter the GC content

is measured by the entire sequences, the third codon position, or the fourfold degenerate sites (Ream et al. 2003).

Subsequently, Musto et al. (2004) published a debate-provoking study. As many environmental factors likely influence genomic GC content evolution, closely related species are expected to differ in fewer environmental factors than distantly related species. The correlation of GC content with growth temperature is less likely to be disturbed by other factors. Therefore, Musto et al. (2004) examined the relationship between genomic GC content and Topt with each prokaryotic family. Among the 20 families they studied, the number of families with positive correlations is significantly higher than expected by chance, no matter the effect of shared ancestry was accounted for or not. Meanwhile, they observed a significant positive correlation when considering all independent contrasts from different families together. However, Marashi and Ghalanbor (2004) noticed that most of the significant correlations within each family depend heavily on the presence of a few outlier species. Exclusion of only one species would lead to loss of significant correlations in several families. Basak et al. (2005) pointed out that the correlation is sensitive to the presence or absence of a few outliers in some families because the sample sizes in these families were too small. Using non-parametric correlation analysis that is not sensitive to the presence of outliers, Musto et al. (2005) repeated their analysis and confirmed their previous results. The debate did not end after that. Wang et al. (2006) updated the Topt values for some species and found that the positive correlation between Topt and genomic GC content in two families disappeared. Besides, they suggested that the positive correlation between Topt and genomic GC content in the family Enterobacteriaceae should be explained by the correlation between genome size and optimal temperature. Still, this study did not shake the confidence of Musto et al. (2006) on the correlation between Topt and genomic GC content in prokaryotes. Although Musto and coauthors have rebutted all the criticisms, their studies have not convinced later authors of review articles (Agashe and Shankar 2014; Meyer 2021). For example, Agashe and Shankar (2014) claimed that "*it seems unlikely that genomic GC content is driven by thermal adaptation*" after reviewing the results of Hurst and Merchant (2001) and Xia et al. (2002), but without mentioning the debates on Musto et al. (2004).

Like the study of Ream et al. (2003) in vertebrates, Zheng and Wu (2010) focused on the GC content in the coding regions of four genes across 815 prokaryotic species, including mesophiles, thermophiles, and hyperthermophiles. They found a positive correlation between the temperature and GC content after controlling a series of environmental and phylogenetic factors. These four genes shared by all the 815 prokaryotic genomes are members of the core genomes.

Using a manually collected dataset of growth temperature and without accounting for the effect of the common ancestor, Sato et al. (2020) recently confirmed the results of Galtier and Lobry (1997). It should be noted that the correlation between Topt and the GC content of structural RNA was consistently observed in much more studies than those mentioned above (Galtier and Lobry 1997; Khachane et al. 2005; Kimura et al. 2006; Kimura et al. 2007; Kimura et al. 2013; Sato et al. 2020).

By contrast, as mentioned above, the correlation between Topt and genomic GC content, if it exists, depends heavily on the sample size, the families of prokaryotes, the particular sequences, and the methods used to detect it.

Benefit from the manually curated dataset of growth temperature from the database TEMPURA (Sato et al. 2020), we carried out a comprehensive analysis on the relationship between growth temperature and GC content. The present study covers three indexes of growth temperature (maximal growth temperature [Tmax], Topt, and minimal growth temperature [Tmin]) and a series of GC content indexes, including GC content of the whole genome (GC_w), GC content of the protein-coding sequences (GC_p), GC content at fourfold degenerate sites (GC_4), GC content of the genes coding structural RNAs (tRNA, GC_{tRNA}; 5S rRNA, GC_{5S}; 16S rRNA, GC_{16S}; 23S rRNA, GC_{23S}) and GC content of non-coding DNA (GC_{non}, including intergenic sequences and untranslated regions of mRNA that are generally unannotated in prokaryotic genomes). The whole genome, primary chromosome genome sequences, plasmid genomes, core genome, and accessory genome of each bacterial or archaeal species have been examined separately. Our results consistently support the existence of a positive correlation between genomic GC content and growth temperature.

Results

Strong phylogenetic signals in both GC contents and growth temperatures

A significant force shaping the prokaryotic evolution is horizontal gene transfer, which makes the genealogical relationships among bacteria and archaea exhibit a somewhat network-like structure. If bifurcation is not the phylogeny's dominant pattern, most phylogenetic comparative methods will not be necessary for prokaryotic evolutionary studies. We are not sure how much this impression has influenced the researchers in prokaryotic genomic studies, but many papers did not use any phylogenetic comparative methods. Despite the frequent horizontal gene transfers, careful examination of the prokaryotic phylogeny could see a statistical tree (Koonin 2015; DeSalle and Riley 2020; Blais and Archibald 2021). In principle, the necessity of phylogenetic comparative methods depends on the significance of the phylogenetic signal, a measure of the correlation between the evolution of the analyzed trait and the presumed phylogenetic tree. We first measured the phylogenetic signals of the analyzed traits for the 681 bacteria and 155 archaea obtained from the database TEMPURA (Sato et al. 2020). As shown in table 1, all the λ values are close to one, which indicates that simple statistical analysis that does not account for common ancestry's effect would lead to inaccurate results (Felsenstein 1985; Symonds and Blomberg 2014).

Bacterial but not archaeal genomic GC contents correlated with growth temperatures

We used the phylogenetic generalized least squares (PGLS) regression to examine the relationships between GC contents and growth temperatures. The significant positive and negative slopes of the

regressions correspond to significant positive and negative correlations, respectively. Four phylogenetic models, the Brownian motion model (BM), the Ornstein-Uhlenbeck model with an ancestral state to be estimated at the root (OUfixedRoot), the Pagel's lambda model (lambda), and the early burst model (EB), have been applied in the analysis. Their results are qualitatively identical and quantitatively similar. Therefore, we present the BM model results in the main text and deposit other models' results as supplementary tables.

Consistent with the numerous previous studies, we found positive correlations between the GC contents of structural RNA genes and growth temperature in bacteria and archaea (Table 2). We noticed a rank in the slope values, from Tmax, Topt, to Tmin.

Unlike previous influential studies (Galtier and Lobry 1997; Hurst and Merchant 2001), we also found positive correlations of Tmax and Topt with various indexes of genomic GC contents, GC_w, GC_p, GC_4, and GC_{non}, in bacteria (Table 2). Nevertheless, bacterial Tmin is not correlated with three of the four GC content indexes (Table 2). In archaea, none of the three temperature indexes (Tmax, Topt, or Tmin) have any significant correlations with any of the four genomic GC content indexes (Table 2).

If growth temperature could shape GC contents by the stabilities of RNA secondary structures and DNA double helix, a structural RNA or a DNA double helix that is stable at the Tmax or Topt is, of course, stable at the Tmin. In this logic, it is reasonable to see that the Tmin has weaker or no significant correlations with GC contents.

The difference in the correlations between bacteria and archaea might be attributed to either unknown intrinsic differences between these two domains or the significant difference in the sample size, 681 *vs.* 155.

Sample sizes matter

If the lack of significant correlations between genomic DNA and Tmax and Topt in archaea results from the small sample size, the correlations in bacteria will be lost when the sample size of bacteria is reduced to 155. For this reason, we randomly selected 155 bacteria from the 681 bacterial samples for 1000 rounds. The results of resampling analysis confirmed the idea, the sample sizes matter (Table 3). In > 950 rounds, the genomic GC content indexes (GC_w, GC_p, GC_4, and GC_{non}) are not correlated with Tmax or Topt ($P > 0.05$). This result could also explain the difference of the present study with Hurst and Merchant (2001), which did not found significant correlations between GC_w/GC_3 and Topt by phylogenetic analysis of about 100 prokaryotes. Meanwhile, a few positive correlation cases happen, indicating that significant positive correlations could also be found by chance when the analyzed sample is small.

Besides, we noticed that the correlations between growth temperature and the GC contents of structural RNA genes might also be lost occasionally when the sample size is severely reduced (Table 3). In the 1000 rounds of resampling, lacking significant correlations happen in 308 (for Tmax) and 473 (for Topt) rounds in 5S rRNA genes, and 12 (for Tmax) and 21 (for Topt) rounds in tRNA genes.

However, in the 16S and 23S rRNA genes, positive correlations were consistently observed in all the 1000 rounds of resampling. We suspected that the tens of times more nucleotides in 16S and 23S rRNA than 5S rRNA make the results of 16S and 23S rRNAs less sensitive to small sample sizes.

In statistics, the rule of thumb boundary between small and large samples is $n = 30$. However, the results in Table 3 indicate that $n = 155$ is a too-small sample in the phylogenetic comparative analyses of the correlation between growth temperature and genomic GC content. Because of the common ancestor, two closely related lineages with highly similar growth temperatures and GC contents should be regarded as nearly one effective sample rather than two independent samples. The effective sample size should be much lower than the census number of the analyzed lineages.

Qualitative data on growth temperature lead to the same conclusion

In the ProTraits database (http://protraits.irb.hr/) and the IMG database (https://img.jgi.doe.gov/)(Brbić et al. 2016; Chen et al. 2020), many prokaryotes lack quantitative measures of growth temperature but are qualitatively classified into four categories: psychrophiles/psychrotrophiles, mesophiles, thermophiles, and hyperthermophiles. We constructed a qualitative dataset of prokaryote growth temperature, including data downloaded from these two datasets and the prokaryotes in the TEMPURA database classified into the four categories referring (Sato et al. 2020). By assigning 1, 2, 3, and 4 to the psychrophiles/psychrotrophiles, mesophiles, thermophiles, and hyperthermophiles, respectively, we transformed the qualitatively characters into numerical values. PGLS regression analyses revealed a positive correlation between GC content and growth temperature in bacteria (slope = 0.457, $P = 0.001$), but not in archaea (slope = –0.582, $P = 0.170$). Although this dataset (4696 bacteria and 279 archaea) is much larger than analyzed above (681 bacteria and 155 archaea), it lost much information during the qualitative classification. All the differences in growth temperature within each category disappear.

We also examined whether the contrast in the temperature category is correlated with the contrast in the GC content between terminal tips of the phylogenetic tree by referring (Aslam et al. 2019). In total, 273 pairs of bacteria and 41 pairs of archaea were obtained from the phylogenetic tree (Parks et al. 2020). Pairwise comparison showed significantly higher GC contents in the bacteria with higher rank in growth temperature (Wilcoxon signed rank test, $P = 0.019$). Still, no significant differences were observed between archaea with different growth temperature ranks (Wilcoxon signed rank test, $P = 0.446$).

Positive correlations observed in genes of both chromosomes and plasmids

Previous studies showed a significant difference in the GC content between plasmids and chromosomes, with significantly lower GC contents in the plasmids (Rocha and Danchin 2002; Nishida 2012; Dietel et al. 2019). Therefore, we examined the correlations between growth temperatures and GC contents separately in chromosomes and plasmids. The separations of plasmids

and chromosomes are arbitrary. We strictly followed the definitions of chromosomes and plasmids of the NCBI genome database (ftp://ftp.ncbi.nlm.nih.gov/genomes/). Among the 681 bacteria and 155 archaea analyzed above, 172 bacteria and 42 archaea have plasmid genomes. The bacterial chromosomes also have GC contents (GC_w, GC_p, GC_4, and GC_{non}) positively correlated with Tmax and Topt (Table 4). Interestingly, the same pattern was also found in the bacterial plasmids (Table 4) in spite that the correlations of Tmax with GC_4 and GC_{non} are just significant at marginal levels ($0.05 < P < 0.1$). All these correlations are not significant in archaea.

In the two previous studies comparing the GC content between plasmids and chromosomes (Rocha and Danchin 2002; Nishida 2012), the common ancestor effect was not accounted for. By the way, we performed a phylogenetic paired t-test (Lindenfors et al. 2010) and confirmed the pattern of lower GC content in plasmids (Table S8).

Positive correlations observed in both core genes and accessory genes

To correspond to the previous gene-centered studies (Ream et al. 2003; 2010), we examined the correlations in bacterial core genes, i.e., genes present in all the bacteria. With the increase in the number of bacteria, the number of core genes decreases rapidly. With a trade-off between the number of core genes and the number of bacteria, we selected 28 core genes present in 420 genomes, mostly ribosomal protein genes. Significant positive correlations have been found between GC contents (GC_p and GC_4) and growth temperatures, Tmax, and Topt (Table 5).

At the opposite side of the core genes, the accessory genes present in one or a few bacteria. When we define the accessory genes as the genes present in less than 5% of the analyzed bacterial genomes, on average, each bacterium has 152 accessory genes. Positive correlations were observed between GC contents (GC_p and GC_4) and growth temperatures (Tmax and Topt), although the values of significance are slightly larger than those in core genes (Table 5). Similar patterns were observed when we increased the threshold in defining accessory genes to 10% ($P < 0.05$ for all cases).

A previous analysis of 36 prokaryotes showed that the core genomes have a significantly higher GC content than accessory genomes (Bohlin et al. 2017). However, this study did not account for the effect of common ancestors. Using a phylogenetic paired t-test (Lindenfors et al. 2010), we did not observe significant differences between the core genes and the accessory genes (Table S15). We also compared the chromosomal accessory genes and plasmid accessory genes using the phylogenetic paired t-test. The accessory genes on chromosomes have significantly higher GC contents than those on plasmids (Table S16).

Discussion

The GC pairs are thermally more stable than AT pairs in both DNA double helix and structural RNAs. However, the difference is not necessarily a strong enough force to shape the evolution of GC content. As RNA structures are more sensitive to the elevation of temperature than DNA double helix, the

growth temperature is stronger in shaping the GC content evolution of the structural RNA genes than in shaping the genomic GC content evolution. Positive correlations between growth temperature and the GC content of structural RNA genes have been repeatedly observed in various prokaryotic studies (Galtier and Lobry 1997; Hurst and Merchant 2001; Khachane et al. 2005; Kimura et al. 2006; Kimura et al. 2007; Kimura et al. 2013; Sato et al. 2020). However, there was a long debate on the correlation between growth temperature and genomic GC content. Benefit from a new manual-curated dataset of prokaryotic growth temperature (Sato et al. 2020), we performed a phylogenetic comparative analysis with a much larger sample than previous studies (Hurst and Merchant 2001; Musto et al. 2004). In 681 bacteria, the genomic GC contents, no matter GC_w, GC_p, GC_4, or GC_{non}, are positively correlated with growth temperatures, Tmax and Topt. However, in 155 archaea, there are no significant correlations. Then, we resampled 155 bacteria from the 682 bacteria for 1000 rounds. In most cases, the significant positive correlations between genomic GC contents and growth temperatures disappeared. The resampling analysis indicates that the small sample sizes of the previous studies might lead to the lack of significant correlations. It is effortless to increase the sample size several times if accurate phylogenetic relationships are not considered in the analysis. As shown in Table 1, we found that both growth temperatures and GC contents exhibit powerful phylogenetic signals. Overlooking the effect of common ancestors would severely affect the accuracy of the results (Felsenstein 1985).

Our resampling analysis indicates that the lack of significant correlations in archaea might result from the small number of effective samples. We hope to repeat the present study in the future with a larger sample of archaeal genomes. However, it should also be kept in mind that the possibility of no correlation between GC content and growth temperature has not been convincingly excluded. Some intrinsic differences between bacteria and archaea might produce a sharp difference in the correlation between GC content and growth temperature. Most prokaryotes have negatively supercoiled DNA, whereas the prokaryotes that grow at temperatures higher than 80°C (mostly archaea) generally have their genomic DNA positively supercoiled with a particular enzyme, reverse gyrase (Vettone et al. 2014). A high level of supercoiling might stabilize the DNA double helix at high temperatures and relieve the high GC content requirement.

A recent study suggests that sequential amino acid substitutions are involved in the thermoadaptation in the archaeal order *Methanococcales* and revealed arginine as the most favoured amino acid (Lecocq et al. 2020). As six GC-rich codons encode the arginine, the thermoadaptation at the proteomic level would affect the evolution of genomic GC content. As the 4-fold degenerate sites are free from the evolutionary forces coming from the natural selection acting on protein sequences, our observations of similar correlations of GC_w, GC_p and GC_4 with growth temperature indicate that the nucleotide composition evolved independently in bacterial adaptation to high temperature.

As the frequent gain and loss of plasmids, the plasmid DNAs could be regarded as accessory genomes. Because of the high turnover rates of plasmids and accessory genes in prokaryotic evolution, we could regard them as new immigrants, as opposed to the natives for the chromosomes and core

296 genes. Although the core genes and even the ribosomal RNA genes may occasionally be transferred
297 across different prokaryotic lineages (Tian et al. 2015; Sato and Miyazaki 2017), the fitness cost of
298 inter-species replacement of homologous sequences (Bershtein et al. 2015) restricts the frequency of
299 the core genes. Genes encoding proteins that carry out essential informational tasks in the cell are less
300 transferable across lineages (Jain et al. 1999; Kacar et al. 2017). Our phylogenetic correlation analysis
301 showed that positive correlations between GC contents and growth temperatures exist in chromosomes
302 and core genes and exist in plasmids and accessory genes. Also, there is no sharp difference in the
303 correlations between the new immigrants and the natives.
304 In large-scale analyses of horizontal gene transfer in prokaryotes, GC-content similarity between
305 donor and recipient was found to be one factor, or one of the factors, governing the compatibility of
306 the new immigrants in new hosts (Popa et al. 2011; Porse et al. 2018). The effect of promoter GC
307 content on the expression of the new immigrants was suggested to be the underlying mechanism
308 governing the compatibility (Gomes et al. 2020). Here, we suggest that the temperature-associated
309 structural stabilities, including the stability of DNA double helix, the stability of the transient DNA-
310 RNA duplex during transcription, and maybe the stability of the possible secondary structures of
311 mature mRNA (Basak et al. 2010), might be another nonexclusive factor governing the compatibility.
312 Previous analyses that did not account for the common ancestor effect showed that chromosomes
313 and core genes have higher GC contents than plasmids and accessory genes, respectively (Rocha and
314 Danchin 2002; Nishida 2012; Bohlin et al. 2017). Our phylogenetic paired t-tests found significant
315 differences between chromosomes and plasmid, but not between core genes and accessory genes.
316 Furthermore, we found that the chromosomal accessory genes have significantly higher GC content
317 than plasmid accessory genes. Bacterial cells usually have multiple copies of plasmids, which coincide
318 with higher maintenance costs (Rocha and Danchin 2002; Dietel et al. 2018; Dietel et al. 2019). We
319 suspected that it is the cost resulting from the high copy number that leads to the GC content
320 difference between the chromosomal accessory genes and plasmid accessory genes. Despite this
321 difference, the GC contents of both plasmids and accessory genes are positively correlated with
322 growth temperature. The new immigrants compatible with the host have a GC content adapted to the
323 host's growth temperature.
324 A previous serial transfer experiment seems to be contradictory to our results. Increased genomic
325 GC content was not observed in the bacterium *P. multocida* after 14,400 generations of increasing
326 temperature from 37°C to 45°C (Xia et al. 2002). Although we observed a positive correlation between
327 genomic GC content and growth temperature, we do not think a small increment in GC content,
328 resulting from either a GC-biased mutator or integration of a GC-rich exogenous sequence, would
329 bring a great advantage to the host organisms. Most likely, it is just a slight advantage. According to
330 the population genetic theory, the slightly beneficial mutants will be efficiently selected only when
331 they are in a large population. The experimental evolution generally involves severe, periodic
332 reductions in population size, and the bottleneck effect dramatically reduces the fixation probability of

beneficial mutations (Wahl et al. 2002). As we see, large-scale statistical analysis has the advantage of revealing slightly beneficial traits.

Finally, we should remark that what we observed are just correlations between GC content and growth temperature, which imply rather than prove the causal effects between the two variables. Besides the thermal adaptation hypothesis (Bernardi and Bernardi 1986), we should be open to other intricate explanations for the observed correlations.

Materials and Methods

We downloaded the prokaryote growth temperatures from the database TEMPURA (Sato et al. 2020). This database contains 8,639 manual curated strains (549 archaea and 8090 bacteria). Using the links to the NCBI Taxonomy database (Federhen 2012) and the taxonomy IDs provided by TEMPURA for each prokaryotic strain, we obtained 1110 prokaryotes whose genome assembly levels were labeled as"complete" from the NCBI database (Sayers et al. 2021). Among them, we found the phylogenetic information for 682 bacteria and 156 archaea from GTDB (Genome Taxonomy Database) (Parks et al. 2020). The sequences of these genomes were downloaded from ftp://ftp.ncbi.nlm.nih.gov/genomes/. To avoid annotation bias resulting from different methods, all the genomes were re-annotated using the DFAST, version 1.2.11, with its default parameters (Tanizawa et al. 2018). In total, we obtained the annotations for 681 bacterial genomes and 155 archaeal genomes. The GC contents of these prokaryotes were calculated from these genome sequences.

We also constructed a large dataset according to their growth temperature qualitatively. First, we divided the 836 prokaryotes mentioned above into four categories according to their growth temperature referring to (Sato et al. 2020): psychrophiles/psychrotrophiles (Topt < 20°C), mesophiles (20 ≤ Topt < 45°C), thermophiles (45 ≤ Topt < 80°C), and hyperthermophiles (80°C ≤ Topt). Then, we downloaded the lists of prokaryotes labeled with psychrophiles/psychrotrophiles, mesophiles, thermophiles, or hyperthermophiles from the ProTraits database (http://protraits.irb.hr/) and the IMG database (https://img.jgi.doe.gov/) (Brbić et al. 2016; Chen et al. 2020). After discarding the overlapping items, the conflicting items, and the items lacking phylogenetic information in the GTDB database (Parks et al. 2020), we obtained a new dataset including 4696 bacteria and 279 archaea (Table S17). The whole-genome GC contents of these prokaryotes were downloaded directly from the NCBI genome database (https://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/prokaryotes.txt).

The phylogenetic signals (λ) of both GC contents and growth temperatures were estimated using the phylosig function of the R (Version 4.0.3) package phytools (Version 0.7-70) (Revell 2012). The PGLS regression was performed using the R (Version 4.0.3) package phylolm (version 2.6.2) (Ho and Ane 2014).

To avoid false-positive results that might happen in multiple correlation analyses of the same dataset, we controlled the false discovery rate by the Benjamini-Hochberg (BH) procedure using the p.adjust function in R (Version 4.0.3).

Supplementary Material

Supplementary data are submitted along with the main text.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number 31671321).

Author Contributions

DKN conceived the study and wrote the manuscript. EZH, XRL, ZLL, and JG performed the data analysis. MYM and DKN All authors read, improved, and approved the final manuscript.

References

Agashe D, Shankar N. 2014. The evolution of bacterial DNA base composition. *J Exp Zool Part B* 322:517-528.

Aslam S, Lan X-R, Zhang B-W, Chen Z-L, Wang L, Niu D-K. 2019. Aerobic prokaryotes do not have higher GC contents than anaerobic prokaryotes, but obligate aerobic prokaryotes have. *BMC Evol Biol* 19:35.

Basak S, Mandal S, Ghosh TC. 2005. Correlations between genomic GC levels and optimal growth temperatures: some comments. *Biochem Biophys Res Commun* 327:969-970.

Basak S, Mukhopadhyay P, Gupta SK, Ghosh TC. 2010. Genomic adaptation of prokaryotic organisms at high temperature. *Bioinformation* 4:352-356.

Bernardi G, Bernardi G. 1986. Compositional constraints and genome evolution. *J Mol Evol* 24:1-11.

Bershtein S, Serohijos AWR, Bhattacharyya S, Manhart M, Choi J-M, Mu W, Zhou J, Shakhnovich EI. 2015. Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria. *PLoS Genet* 11:e1005612.

Blais C, Archibald JM. 2021. The past, present and future of the tree of life. *Curr Biol* 31:R314-R321.

Bohlin J, Eldholm V, Pettersson JHO, Brynildsrud O, Snipen L. 2017. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genomics* 18:151.

Brbić M, Piškorec M, Vidulin V, Kriško A, Šmuc T, Supek F. 2016. The landscape of microbial phenotypic traits and associated genes. *Nucleic Acids Res* 44:10074-10090.

Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter S, Varghese N, Seshadri R, et al. 2020. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res* 49:D751-D763.

DeSalle R, Riley M. 2020. Should networks supplant tree building? *Microorganisms* 8.

Dietel A-K, Merker H, Kaltenpoth M, Kost C. 2019. Selective advantages favour high genomic AT-contents in intracellular elements. *PLoS Genet* 15:e1007778.

Dietel AK, Kaltenpoth M, Kost C. 2018. Convergent evolution in intracellular elements

411 plasmids as model endosymbionts. *Trends Microbiol* 26:755-768.
412 Federhen S. 2012. The NCBI Taxonomy database. *Nucleic Acids Res* 40:D136-D143.
413 Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat* 125:1-15.
414 Foerstner KU, von Mering C, Hooper SD, Bork P. 2005. Environments shape the nucleotide
415 composition of genomes. *EMBO Rep* 6:1208-1213.
416 Galtier N, Lobry JR. 1997. Relationships between genomic G+C content, RNA secondary
417 structures, and optimal growth temperature in prokaryotes. *J Mol Evol* 44:632-636.
418 Glemin S, Clement Y, David J, Ressayre A. 2014. GC content evolution in coding regions of
419 angiosperm genomes: a unifying hypothesis. *Trends Genet* 30:263-270.
420 Gomes ALC, Johns NI, Yang A, Velez-Cortes F, Smillie CS, Smith MB, Alm EJ, Wang HH.
421 2020. Genome and sequence determinants governing the expression of horizontally acquired
422 DNA in bacteria. *ISME J* 14:2347-2357.
423 Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content
424 in bacteria. *PLoS Genet* 6:e1001107.
425 Ho LST, Ane C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution
426 models. *Syst Biol* 63:397-408.
427 Hurst LD, Merchant AR. 2001. High guanine-cytosine content is not an adaptation to high
428 temperature: a comparative analysis amongst prokaryotes. *Proc R Soc B* 268:493-497.
429 Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: The complexity
430 hypothesis. *Proc Natl Acad Sci USA* 96:3801-3806.
431 Kacar B, Garmendia E, Tuncbag N, Andersson DI, Hughes D. 2017. Functional constraints on
432 replacing an essential gene with its ancient and modern homologs. *mBio* 8:e01276-01217.
433 Kagawa Y, Nojima H, Nukiwa N, Ishizuka M, Nakajima T, Yasuhara T, Tanaka T, Oshima T.
434 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile.
435 DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *The Journal*
436 *of biological chemistry* 259:2956-2960.
437 Khachane AN, Timmis KN, dos Santos VAPM. 2005. Uracil content of 16S rRNA of
438 thermophilic and psychrophilic prokaryotes correlates inversely with their optimal growth
439 temperatures. *Nucleic Acids Res* 33:4016-4022.
440 Kimura H, Ishibashi J-I, Masuda H, Kato K, Hanada S. 2007. Selective phylogenetic analysis
441 targeting 16S rRNA genes of hyperthermophilic archaea in the deep-subsurface hot biosphere.
442 *Appl Environ Microbiol* 73:2110-2117.
443 Kimura H, Mori K, Yamanaka T, Ishibashi J-I. 2013. Growth temperatures of archaeal
444 communities can be estimated from the guanine-plus-cytosine contents of 16S rRNA gene
445 fragments. *Env Microbiol Rep* 5:468-474.
446 Kimura H, Sugihara M, Kato K, Hanada S. 2006. Selective phylogenetic analysis targeted at
447 16S rRNA genes of thermophiles and hyperthermophiles in deep-subsurface geothermal
448 environments. *Appl Environ Microbiol* 72:21-27.
449 Koonin EV. 2015. The turbulent network dynamics of microbial evolution and the statistical
450 tree of life. *J Mol Evol* 80:244-250.
451 Lecocq M, Groussin M, Gouy M, Brochier-Armanet C. 2020. The molecular determinants of
452 thermoadaptation: *Methanococcales* as a case study. *Mol Biol Evol*.
453 Lindenfors P, Revell LJ, Nunn CL. 2010. Sexual dimorphism in primate aerobic capacity: a
454 phylogenetic test. *J Evol Biol* 23:1183-1194.

Mann S, Chen YPP. 2010. Bacterial genomic G plus C composition-eliciting environmental adaptation. *Genomics* 95:7-15.

Marashi S-A, Ghalanbor Z. 2004. Correlations between genomic GC levels and optimal growth temperatures are not 'robust'. *Biochem Biophys Res Commun* 325:381-383.

Meyer MM. 2021. Revisiting the relationships between genomic G plus C content, RNA secondary structures, and optimal growth temperature. *J Mol Evol* 89:165-171.

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valin F, Bernardi G. 2005. The correlation between genomic G+C and optimal growth temperature of prokaryotes is robust: A reply to Marashi and Ghalanbor. *Biochem Biophys Res Commun* 330:357-360.

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valin F, Bernardi G. 2004. Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Lett* 573:73-77.

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valin F, Bernardi G. 2006. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem Biophys Res Commun* 347:1-3.

Nguyen DT, Wu B, Xiao S, Hao W. 2020. Evolution of a record-setting AT-rich genome: indel mutation, recombination, and substitution bias. *Genome Biol Evol* 12:2344-2354.

Nishida H. 2012. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *Int J Evol Biol* 2012:342482.

Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38:1079-1086.

Popa O, Hazkani-Covo E, Landan G, Martin W, Dagan T. 2011. Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res* 21:599-609.

Porse A, Schou TS, Munck C, Ellabaan MMH, Sommer MOA. 2018. Biochemical mechanisms determine the functional compatibility of heterologous genes. *Nat Commun* 9:522.

Raghavan R, Kelkar YD, Ochman H. 2012. A selective force favoring increased G+C content in bacterial genes. *Proc Natl Acad Sci USA* 109:14504-14507.

Ream RA, Johns GC, Somero GN. 2003. Base compositions of genes encoding alpha-actin and lactate dehydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G+C content. *Mol Biol Evol* 20:105-110.

Reichenberger ER, Rosen G, Hershberg U, Hershberg R. 2015. Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. *Genome Biol Evol* 7:1380-1389.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217-223.

Rocha EPC, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. *Trends Genet* 18:291-294.

Sato M, Miyazaki K. 2017. Phylogenetic network analysis revealed the occurrence of horizontal gene transfer of 16S rRNA in the genus *Enterobacter*. *Front Microbiol* 8:10.

Sato Y, Okano K, Kimura H, Honda K. 2020. TEMPURA: database of growth TEMPeratures of Usual and RAre Prokaryotes. *Microbes and Environments* 35:ME20074.

Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, et al. 2021. Database resources of the National Center for Biotechnology

Information. *Nucleic Acids Res* 49:D10-D17.

Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, Tichý L, Grulich V, Rotreklová O. 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci USA* 111:E4096-E4102.

Symonds MRE, Blomberg SP. 2014. A primer on phylogenetic generalised least squares. In: Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 105-130.

Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34:1037-1039.

Tian R-M, Cai L, Zhang W-P, Cao H-L, Qian P-Y. 2015. Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene. *Genome Biol Evol* 7:2310-2320.

Vettone A, Perugino G, Rossi M, Valenti A, Ciaramella M. 2014. Genome stability: recent insights in the topoisomerase reverse gyrase and thermophilic DNA alkyltransferase. *Extremophiles* 18:895-904.

Wahl LM, Gerrish PJ, Saika-Voivod I. 2002. Evaluating the impact of population bottlenecks in experimental evolution. *Genetics* 162:961-971.

Wang H-C, Susko E, Roger AJ. 2006. On the correlation between genomic G+C content and optimal growth temperature in prokaryotes: Data quality and confounding factors. *Biochem Biophys Res Commun* 342:681-684.

Weissman JL, Fagan WF, Johnson PLF. 2019. Linking high GC content to the repair of double strand breaks in prokaryotic genomes. *PLoS Genet* 15:e1008493.

Winter G, Koch GLE, Hartley BS, Barker DG. 1983. The amino acid sequence of the tyrosy1-tRNA synthetase from *Bacillus stearothermophilus*. *Eur J Biochem* 132:383-387.

Wu H, Zhang Z, Hu S, Yu J. 2012. On the molecular mechanism of GC content variation among eubacterial genomes. *Biol Direct* 7:2.

Xia X, Wei T, Xie Z, Danchin A. 2002. Genomic changes in nucleotide and dinucleotide frequencies in *Pasteurella multocida* cultured under high temperature. *Genetics* 161:1385-1394.

Zheng H, Wu H. 2010. Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. *BMC Bioinformatics* 11:S7.

Table 1. The phylogenetic signals of the variables analyzed in this study.

Traits	Bacteria			Archaea		
	n	Pagel′s λ	*P*	*n*	Pagel′s λ	*P*
Tmax	681	0.957	3.5×10^{-178}	155	1.000	1.5×10^{-72}
Topt,	681	0.950	1.1×10^{-196}	155	0.988	5.7×10^{-70}
Tmin	681	0.933	6.6×10^{-152}	155	0.966	1.9×10^{-53}
GC_w	681	1.000	2.6×10^{-294}	155	1.000	5.8×10^{-60}
GC_p	681	1.000	4.7×10^{-292}	155	1.000	2.4×10^{-59}
GC_4	681	1.000	4.8×10^{-238}	155	1.000	5.1×10^{-53}
GC_{non}	681	1.000	8.6×10^{-305}	155	1.000	2.0×10^{-65}
GC_{tRNA}	681	0.998	6.0×10^{-275}	155	1.000	2.1×10^{-91}
GC_{5S}	646	1.000	7.0×10^{-178}	130	1.000	9.1×10^{-51}
GC_{16S}	681	0.999	8.7×10^{-250}	155	0.996	3.7×10^{-86}
GC_{23S}	681	1.000	2.1×10^{-245}	155	1.000	6.8×10^{-83}

Tmax, Topt, and Tmin represent maximal, optimal, and minimal growth temperature, respectively; GC_w, GC_p, GC_4, GC_{tRNA}, GC_{5S}, GC_{16S}, GC_{23S}, and GC_{non} represent the GC contents of the genome, the protein-coding sequences, the fourfold degenerate sites, the genes coding tRNAs, the genes coding 5S rRNA, the genes coding 16S rRNA, the genes coding 23S rRNA, and the non-coding DNA (including intergenic sequences and untranslated regions of mRNA), respectively. The phylogenetic signals of the chromosomal genes, the plasmid genes, the core genes, and the accessory genes are also very close to one and deposited in supplementary Table S1-S4.

544 Table 2. PGLS analysis of GC contents and growth temperatures.

	Bacteria			Archaea		
	Slope	P	P_{BH}	Slope	P	P_{BH}
GC_w-Tmax	7.1×10^{-4}	7.1×10^{-4}	9.4×10^{-4}	6.6×10^{-4}	0.115	0.153
GC_w-Topt	5.7×10^{-4}	0.009	0.011	3.3×10^{-4}	0.377	0.503
GC_w-Tmin	2.8×10^{-4}	0.156	0.226	5.2×10^{-4}	0.126	0.168
GC_p-Tmax	6.6×10^{-4}	0.002	0.002	5.6×10^{-4}	0.183	0.209
GC_p-Topt	5.3×10^{-4}	0.015	0.016	2.4×10^{-4}	0.522	0.597
GC_p-Tmin	2.5×10^{-4}	0.202	0.231	4.6×10^{-4}	0.180	0.205
GC_4-Tmax	0.001	0.003	0.003	9.9×10^{-4}	0.321	0.321
GC_4-Topt	0.001	0.016	0.016	2.2×10^{-4}	0.806	0.806
GC_4-Tmin	5.5×10^{-4}	0.239	0.239	6.9×10^{-4}	0.393	0.393
GC_{non}-Tmax	8.0×10^{-4}	1.7×10^{-4}	2.8×10^{-4}	9.1×10^{-4}	0.025	0.041
GC_{non}-Topt	6.4×10^{-4}	0.004	0.006	6.4×10^{-4}	0.080	0.129
GC_{non}-Tmin	2.7×10^{-4}	0.170	0.226	6.5×10^{-4}	0.048	0.077
GC_{tRNA}-Tmax	4.1×10^{-4}	2.2×10^{-16}	5.9×10^{-16}	7.1×10^{-4}	1.8×10^{-11}	7.2×10^{-11}
GC_{tRNA}-Topt	3.9×10^{-4}	2.6×10^{-14}	6.9×10^{-14}	5.0×10^{-4}	2.5×10^{-7}	6.7×10^{-7}
GC_{tRNA}-Tmin	1.5×10^{-4}	9.1×10^{-4}	0.002	4.2×10^{-4}	1.8×10^{-6}	4.7×10^{-6}
GC_{5S}-Tmax	5.5×10^{-4}	1.2×10^{-6}	2.4×10^{-6}	0.001	1.9×10^{-5}	3.9×10^{-5}
GC_{5S}-Topt	4.4×10^{-4}	1.4×10^{-4}	2.9×10^{-4}	8.9×10^{-4}	1.6×10^{-4}	3.2×10^{-4}
GC_{5S}-Tmin	3.5×10^{-4}	0.001	0.002	6.1×10^{-4}	0.005	0.010
GC_{16S}-Tmax	5.4×10^{-4}	2.2×10^{-16}	5.9×10^{-16}	8.2×10^{-4}	3.9×10^{-11}	1.0×10^{-10}
GC_{16S}-Topt	5.2×10^{-4}	2.2×10^{-16}	8.8×10^{-16}	7.2×10^{-4}	1.1×10^{-10}	4.5×10^{-10}
GC_{16S}-Tmin	4.6×10^{-4}	2.2×10^{-16}	8.8×10^{-16}	5.5×10^{-4}	8.5×10^{-8}	3.4×10^{-7}
GC_{23S}-Tmax	6.6×10^{-4}	2.2×10^{-16}	5.9×10^{-16}	0.001	2.2×10^{-16}	1.8×10^{-15}
GC_{23S}-Topt	6.5×10^{-4}	2.2×10^{-16}	8.8×10^{-16}	0.001	1.2×10^{-14}	9.5×10^{-14}
GC_{23S}-Tmin	4.9×10^{-4}	2.2×10^{-16}	8.8×10^{-16}	8.3×10^{-4}	8.0×10^{-11}	6.4×10^{-10}

545 GC contents were the dependent variables, and growth temperatures were the independent variables.
546 The results in this table were obtained using the Brownian motion model. Similar results obtained
547 from three other models are deposited in supplementary Table S5-S7. P_{BH}, Benjamini-Hochberg
548 adjusted P value. Please see Table 1 for the meanings of the other abbreviations.

549

550 Table 3. The appearance of correlations in 1000 rounds of random resampling 155 samples from the
551 681 bacteria.

	Significant negative ($P < 0.05$)	Not significant ($P > 0.05$)	Significant positive ($P < 0.05$)
GC_w-Tmax	0	974	26
GC_w-Topt	0	991	9
GC_p-Tmax	0	976	24
GC_p-Topt	0	993	7
GC_4-Tmax	0	962	38
GC_4-Topt	0	992	8
GC_{non}-Tmax	0	974	26
GC_{non}-Topt	0	992	8
GC_{tRNA}-Tmax	0	12	988
GC_{tRNA}-Topt	0	21	979
GC_{5S}-Tmax	0	308	692
GC_{5S}-Topt	0	473	527
GC_{16S}-Tmax	0	0	1000
GC_{16S}-Topt	0	0	1000
GC_{23S}-Tmax	0	0	1000
GC_{23S}-Topt	0	0	1000

552 GC contents were the dependent variables, and growth temperatures were the independent variables.
553 The results in this table were obtained using the Brownian motion model. Please see Table 1 for the
554 meanings of the other abbreviations. The datasets for each round of resampling are deposited in
555 Supplementary Data S1.

556

557

Table 4. PGLS analysis of GC contents and growth temperatures in chromosomes and plasmids.

	Plasmid			Chromosome		
	Slope	*P*	P_{BH}	Slope	*P*	P_{BH}
GC_w-Tmax	0.001	0.009	0.043	9.6×10^{-4}	0.029	0.043
GC_w-Topt	0.001	0.005	0.031	9.6×10^{-4}	0.023	0.031
GC_p-Tmax	0.001	0.016	0.043	9.1×10^{-4}	0.038	0.046
GC_p-Topt	0.001	0.010	0.031	9.2×10^{-4}	0.031	0.034
GC_4-Tmax	0.002	0.072	0.072	0.002	0.027	0.043
GC_4-Topt	0.002	0.044	0.044	0.002	0.017	0.031
GC_{non}-Tmax	8.3×10^{-4}	0.055	0.060	0.001	0.021	0.043
GC_{non}-Topt	9.3×10^{-4}	0.025	0.031	0.001	0.021	0.031

GC contents were the dependent variables, and growth temperatures were the independent variables. The results in this table were obtained using the Brownian motion model. Similar results obtained from three other models are deposited in supplementary Table S9-S11. P_{BH}, Benjamini-Hochberg adjusted *P* value. Please see Table 1 for the meanings of the other abbreviations.

Table 5. PGLS analysis of GC contents and growth temperatures in both core genes and accessory genes

	Core genes			Accessory genes		
	Slope	*P*	P_{BH}	Slope	*P*	P_{BH}
GC_p-Tmax	7.6×10^{-4}	9.6×10^{-4}	0.002	9.0×10^{-4}	0.001	0.002
GC_p-Topt	6.4×10^{-4}	0.007	0.025	6.3×10^{-4}	0.026	0.030
GC_4-Tmax	0.002	6.3×10^{-4}	0.002	0.002	0.003	0.003
GC_4-Topt	0.002	0.004	0.025	0.002	0.019	0.030

GC contents were the dependent variables, and growth temperatures were the independent variables. The results in this table were obtained using the Brownian motion model. Similar results obtained from three other models are deposited in supplementary Table S12-S14. P_{BH}, Benjamini-Hochberg adjusted *P* value. Please see Table 1 for the meanings of the other abbreviations.