Opinion

Genomic and proteomic adaptations to growth at high temperature
Donal A Hickey* and Gregory AC Singer†

Addresses: *Department of Biology, Concordia University, 7141 Sherbrooke Street, Montreal, Quebec, H4B 1R6, Canada. †Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH 43210, USA.

Correspondence: Donal A Hickey. E-mail: dhickey@alcor.concordia.ca

Abstract

Most positively selected mutations cause changes in metabolism, resulting in a better-adapted phenotype. But as well as acting on the information content of genes, natural selection may also act directly on nucleic acid and protein molecules. We review the evidence for direct temperature-dependent natural selection acting on genomes, transcriptomes and proteomes.

What's so special about adaptation to growth at high temperature?

Variations in environmental temperature represent an obvious and easily quantifiable form of environmental hetero-geneity. Biologists have long been aware of a host of behavioral, morphological and physiological adaptations to this environmental variable. Recently, the accumulation of genomic data has led to an interest in another type of temperature adaptation. Specifically, we would like to know whether the genomes themselves - along with their encoded proteomes - are subject to predictable, temperature-dependent patterns of molecular evolution.

While variations in environmental temperature share many of the characteristics of other environmental variables, temperature is special because of its pervasiveness: it can penetrate physical barriers and can have dramatic effects on the structure of virtually all macromolecules. And given that temperature variation affects all levels of biological adaptation, we see adaptive responses at all of these levels. For instance, variations in environmental temperature can be used to explain the evolution of biological phenomena as diverse as the migration patterns of birds, on the one hand, or the density of hydrogen bonds in a nucleic acid sequence, on the other.

Adaptations at the genome (DNA) level

Ever since the experimental demonstration that the thermal denaturation of double-stranded DNA molecules is affected by their nucleotide composition [1], biologists have been intrigued by the possibility that the same principles would apply in nature. The expectation (which is both perfectly logical and supported by laboratory experiments) is that the genomes of organisms growing at higher temperature would be subject to selection for a higher proportion of G+C than A+T, because of the increased number of hydrogen bonds between G and C than A and T on complementary strands. Despite some early reports of supporting evidence based on single gene sequences, however, more extensive sequencing of entire bacterial genomes shows quite convincingly, although unexpectedly, that there is no obvious correlation between the G+C content of the genome and the optimal environmental growth temperature of the organism [2-4]. Indeed, many highly thermophilic species, such as *Pyrococcus abyssi* and *Aquifex aeolicus*, have genomic G+C contents of less than 50%, while some mesophiles - such as the human parasite *Mycobacterium tuberculosis* - have much higher G+C contents in their genomes. It appears that the large variations in the average genomic G+C content between species are largely the result of biased mutation and repair pressures [5-10]. We must conclude that thermophiles

have mechanisms other than increasing G+C content for maintaining the double-stranded structure of their DNA at high temperatures (Figure 1). Two possibilities are the existence of thermophile-specific enzymes, such as the reverse gyrase [11], or selection for certain dinucleotides that may contribute to thermostability [12].

A number of recent studies (discussed in more detail below) have shown other sequence differences between mesophiles and thermophiles, such as the increased level of purine bases in the coding strands of thermophiles [4,8,13,14]. While these effects can be detected at the DNA level, and may be due to the effects of natural selection, they reflect selection for RNA stability rather than direct selection on DNA.

Adaptations at the transcriptome (RNA) level

The transcriptome includes both the structural RNAs (such as ribosomal and transfer RNAs, rRNAs and tRNAs) and the protein-encoding messenger RNAs. One could argue that these molecules, especially the structural RNAs, would be subject to the same temperature-dependent constraints as DNA. Of course, given that the expected correlation between G+C content of genomic DNA and growth temperature is not seen, we might expect that the correlation would also be lacking at the RNA level. But, interestingly, this is not the case. For instance, Galtier and Lobry [2] demonstrated that there is a significant correlation between the G+C content of structural RNAs and growth temperature, and that the high G+C content was concentrated in the double-stranded stem regions of the molecule. This provides strong evidence for selection acting to increase the thermostability of these

Figure 1
Selection for growth at high temperature affects many molecular processes simultaneously.

regions by changing the nucleotide composition. Indeed, this enrichment of G and C is so striking that structural RNA genes virtually identify themselves within the genomes of hyperthermophiles whose DNA is otherwise AT-rich [15].

The effects of natural selection are not limited to the double-stranded regions of these RNAs, however: selection is also acting to reduce the G+C content of the single-stranded regions of rRNA molecules, thus maintaining them in the single-stranded state [13]. An obvious question that comes to mind is why we observe the expected correlation between nucleotide content and growth temperature in the paired regions of an RNA molecule, but not in double-stranded DNA. One possible answer is that single mutations affecting nucleotide composition have a much greater effect on the stability of the stem regions of an RNA molecule than they do on double-stranded genomic DNA, simply because the length of the paired region is much shorter in the RNA molecule.

In contrast to structural RNAs, the critical feature of the protein-coding messenger RNAs is not their secondary structure but their coding capacity. Thus we might not *a priori* expect to see strong selection for structural stability in these molecules. While it is true that a given, specific secondary structure may not be important for mRNAs, stability *per se* is critically important, because it affects the steady-state level of the genetic message within the cell. There is now growing evidence [8,13,14,16-18] that all single-stranded RNA molecules, along with the single-stranded segments of structural RNAs, show characteristic patterns of nucleotide composition in all organisms. Specifically, they are relatively rich in purines, particularly adenine [13,14,16]. Moreover, the degree of purine-richness correlates with environmental growth temperature. The initial interpretation of these trends [17] was that they acted to prevent purine-pyrimidine base pairing between coding sequences. Such base pairing would be prevented by having a preponderance of one type of base - either purines or pyrimidines - on the coding strand. Subsequent studies [4,8,13] indicate, however, that the selection is specifically for purines.

Translational efficiency and codon usage at high temperature

Although different synonymous codons may encode a single amino acid, there has been considerable interest in the possibility that some codons are functionally 'preferred'. The idea of preferred codons stems from the work of Ikemura [19], who showed a positive correlation between the frequency of particular codons and the abundance of their cognate tRNAs. Over the past two decades, many genomic studies have attempted to detect clear evidence for selection acting on synonymous codons, but despite all of these studies it now appears that the major determinant of synonymous codon usage on a genome-wide scale is mutational bias rather than selection [10,20-22]. Despite the dominant effect of nucleotide composition, recent genomic surveys have shown
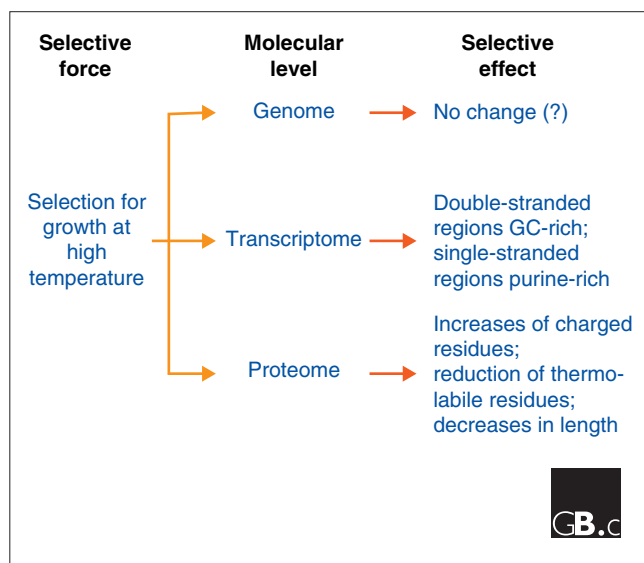
that environmental growth temperature can have an important secondary effect on patterns of synonymous codon usage [8,23,24]. Although there is no obvious explanation for why particular codons are used preferentially among thermophiles, the fact that the pattern is repeated within different evolutionary lineages provides strong support for the fact that it is based on natural selection.

Adaptations at the proteome level

Given that the thermolability of protein structures - like that of nucleic acid structures - can easily be demonstrated in the laboratory, and since protein function depends on protein structure, we expect the proteins of thermophilic organisms to have been subjected to intense natural selection for stability at high temperature. It is, however, difficult to predict the precise outcome of such selection because the forces governing protein structure and function are not yet well understood. Many comparisons of individual protein sequences between mesophiles and thermophiles have been reported in the recent literature. Although several of these studies point to differences between thermophilic proteins and their mesophilic homologs, different studies have tended to identify different aspects of protein sequence and structure as contributing to thermostability [25]. The attraction of studying entire proteomes is that we can hope to identify the more 'universal' adaptations underlying protein stability at high temperature. But, as pointed out by Petsko [26], the problem with such genome-wide studies is that they may only discover some of the lowest common denominators for thermal adaptation at the protein level.

Most of the proteome-based studies to date have focused on the average amino-acid composition of proteins in the proteomes of mesophiles and thermophiles. If we consider that protein structure is determined to a large extent by the primary amino-acid sequences, then we can look for consistent differences in amino-acid composition between the proteins of thermophiles and mesophiles. Such differences have been reported for individual genes and in whole-genome comparisons [8,27-29]. These studies show that while the average amino-acid composition of a given proteome is dramatically affected by the underlying patterns of genomic nucleotide bias [6,9], there is a secondary but highly significant effect of growth temperature. One study [21] found a significant effect of nucleotide bias, but did not reveal any selection on the amino-acid content of thermophilic proteins. By limiting the analysis to a subset of genomes with comparable nucleotide compositions, we [8] showed that the major effect of thermophily at the proteome level was a significant reduction in the frequency of the thermolabile amino acids histidine, glutamine and threonine. This is consistent with the recent observation of increased evolutionary constraint on thermophilic proteomes [30]. The concomitant increase, among thermophiles, of both positively charged residues (arginine and lysine) and negatively

charged residues (glutamic acid) suggests that ionic bonds between oppositely charged residues may help to stabilize multimeric proteins at high temperature [28]. The proteomes of thermophiles also contain a larger fraction of proteins with isoelectric points in the basic range [31], and a general bias in favor of charged rather than polar residues among thermophiles has been noted in two separate studies [32,33]. One of the genome-wide surveys [28] also found support for the conclusions of previous pilot studies (based on one or a few genes) that there are average length differences between the proteins of mesophilic and thermophilic species [32,34,35]. Specifically, the proteins of thermophiles tend to be somewhat shorter than their mesophilic homologs. Finally, a number of recent structural genomics studies [36-39] support the sequence-based studies in that they point to an increase in intra-helical salt bridges and in hydrogen-bond formation among thermophiles. The increased number of salt bridges may contribute to protein stability at high temperature [40].

Post-translational molecular adaptations in thermophiles

Most species can survive for short periods of time at temperatures that are significantly higher than their normal growth temperature. Such a pulse of increased temperature usually triggers the expression of heat-shock proteins that act as chaperones to facilitate protein stabilization and proper protein folding. Such protein chaperones do, in fact, also play a role in thermophiles [41]. Furthermore, genome-sequence surveys have uncovered evidence for a novel, thermophile-specific set molecular chaperones among highly thermophilic species [42]. Thus, in addition to encoding more thermostable mRNAs and proteins, thermophilic organisms may devote more energy to the stabilization of those proteins at high temperature.

Complications of genome-wide surveys
Secondary effects of selection

A significant complication in genomic surveys, although one that is often ignored, is that the average patterns seen in genomes and proteomes are not independent; for instance, the nucleotide composition of the genome can have a dramatic effect on the amino-acid composition of the encoded proteome [6,43,44]. Although most of the studies to date have looked at the effect of G+C content on protein composition, similar effects will result from other kinds of genomic biases [45,46]. For instance, a genome whose coding regions are very rich in purines will necessarily encode a proteome that is deficient in phenylalanine residues, and a genome with pyrimidine-rich coding regions would correspondingly encode few lysines and glutamic acids. Thus, if the sequences on the coding strand are subject to selection for increased purine content because of increased mRNA stability, this selection at the level of RNA can result in a correlated change in the amino-acid content

of the proteins, and even in deterministic changes in the biochemical properties of these proteins - the isoelectric point, for example. Many recent studies have discussed the possibility that mutational biases can mimic the effects of selection, but few authors seem aware of the problem where a selective effect at one level results in an apparent selective effect at another level.

The need for replication

Large-scale genomic comparisons include, by definition, a large amount of information. Typically, thousands of genes are scored and this can give the impression of ample replication, leading to high statistical confidence in the results. In many genomic comparisons, however, although very many gene sequences are included in the analysis, as few as two genomes may be considered. Any systemic bias in the data that may occur within a given genome is not corrected by sampling more genes from the same source; in fact, the inclusion of more genes simply enhances the problem [47,48]. Not only do we need to replicate our observations over many genomes, but we also need to be aware that those genomes are not independent samples because of their phylogenetic relationships. For instance, if we compare several thermophilic species, all of which happen to be archaea, with several mesophiles, all of which are eubacteria, we cannot tell if the differences that we observe are due to the effects of natural selection acting independently on many genes and genomes, or due to a single event that occurred early in the phylogenetic history of the two groups (Figure 2). We must be able to demonstrate that a given evolutionary solution for growth at high temperatures can cross phylogenetic boundaries - that it can arise more than once in the phylogenetic tree of the genomes under study. Using this approach, Musto *et al.* [49] have recently uncovered evidence in favor of a correlation between genomic GC content and optimal growth temperature.

What about thermophilic eukaryotes?

The ability to grow at high temperature is relatively common among archaeal species, and several thermophilic species of eubacteria have also been described. Among the eukaryotes, however, thermophily is much rarer [50] and there are no hyperthermophiles among the eukaryotes. The upper limit for thermophilic eukaryotes is approximately 60°C [51]. Even at this relatively modest temperature (relative to those tolerated by thermophilic prokaryotes), we do not find any complex, multicellular eukaryotes. It has been suggested that eukaryotes are not thermophilic because of the susceptibility of their mRNA to degradation at high temperature [52], and growth at very high temperatures may also require the presence of special lipids that are not found in eukaryotes [53]. While these constraints apply to all eukaryotes, for multicellular animals the temperature threshold is not set at the molecular level but at the physiological level. Specifically, increasing oxygen demand at
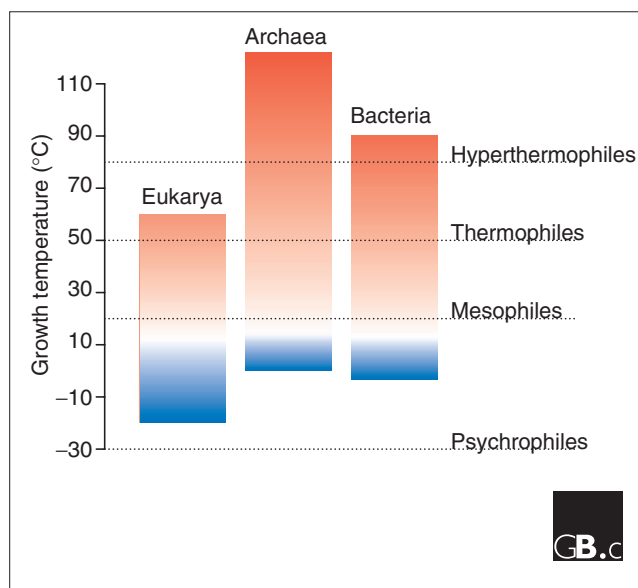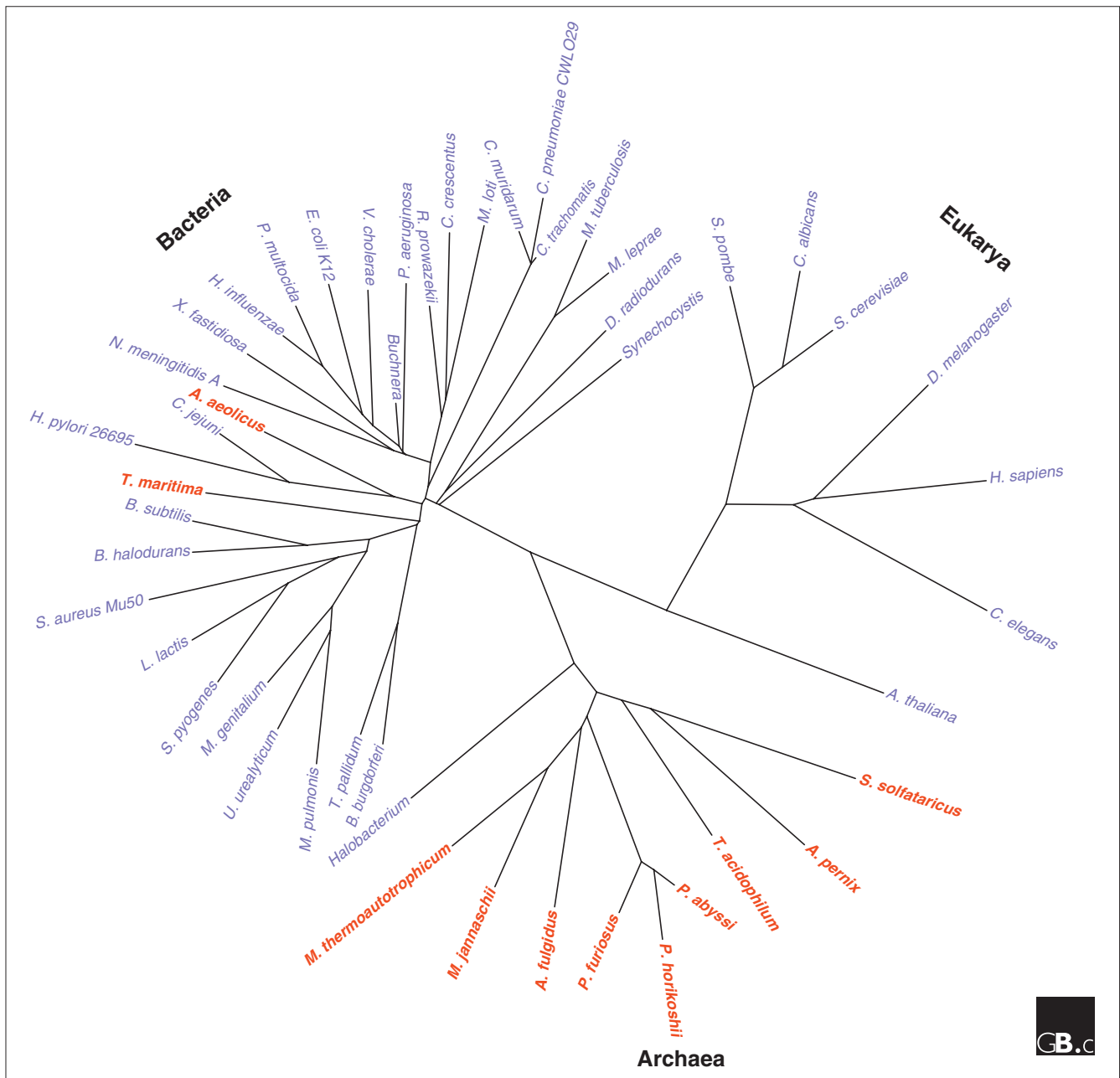

Figure 2
The phylogenetic distribution of thermophily. The ability to grow at high temperature is common among the archaea, relatively rare among eubacteria, and virtually absent among eukaryotes. The growth temperatures were taken from the Prokaryotic Growth Temperature Database [61].

higher temperatures results in depleted oxygen levels in the body fluids [54]. This explains why multicellular animals are even more restricted in their temperature ranges than are microbial eukaryotes (Figures 2 and 3).

Several authors have drawn parallels between thermophilic and mesophilic microbes on the one hand, and warm- and cold-blooded vertebrates on the other. In fact, a considerable amount of work has been done on the correlation of differences in genomic G+C content with the body temperature of animals [55]. Although at first glance there does appear to be a convincing correlation between elevated genomic G+C content (especially in isochore regions) and homeothermy, these results are subject to alternative explanations. For instance, the higher G+C content in certain regions of mammalian genomes may be due to elevated recombination rates in those regions [56,57]. It is also worth noting that the body temperature of mammals is well below 45 °C, which is usually taken as the lower threshold for thermophily among prokaryotes.

In conclusion, given that temperature is a single, clearly defined environmental variable, one might expect to see a single, characteristic genomic and/or proteomic response to changes in this variable. We do see selective responses at the nucleic acid and protein levels, but they are varied and unpredictable. It is especially difficult to predict any significant differences above the level of primary sequence composition.

Figure 3
Temperature tolerance ranges of species of eubacteria, eukaryotes and archaea, illustrated on a phylogenetic tree using the SHOT web server [62]. Species that grow at temperatures above 50°C are indicated in red; the remaining species grow below 50°C. Eukaryotes have a much lower thermal tolerance than either archaea or eubacteria. The following species have been used: *Aeropyrum pernix*, *Aquifex aeolicus*, *Arabidopsis thaliana*, *Archaeoglobus fulgidus*, *Bacillus halodurans*, *Bacillus subtilis*, *Borrelia burgdorferi*, *Buchnera sp.*, *Caenorhabditis elegans*, *Campylobacter jejuni*, *Candida albicans*, *Caulobacter crescentus*, *Chlamydia muridarum*, *Chlamydia trachomatis*, *Chlamydophila pneumoniae* CWL029, *Deinococcus radiodurans*, *Drosophila melanogaster*, *Escherichia coli* K12, *Haemophilus influenzae*, *Halobacterium salinarum*, *Helicobacter pylori* 26695, *Homo sapiens*, *Leuconostoc lactis*, *Mesorhizobium loti*, *Methanocaldococcus jannaschii*, *Methanobacter thermoautotrophicum*, *Methanosaeta thermophila*, *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycoplasma genitalium*, *Mycoplasma pulmonis*, *Neisseria meningitidis* A, *Pasteurella multocida*, *Pseudomonas aeruginosa*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Rickettsia prowazekii*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Sulfolobus solfataricus*, *Synechocystis sp.* PCC6803, *Thermoplasma acidophilum*, *Thermotoga maritima*, *Treponema pallidum*, *Ureaplasma urealyticum*, *Vibrio cholerae*, and *Xylella fastidiosa*.

A number of general trends have been identified in the sequence composition of DNA, RNA and proteins, but it has proved much more difficult to identify thermophilic responses at the higher levels of structural organization. This is particularly true of protein structure, partly because we do not yet have a good understanding of the rules governing

protein folding, and partly because it now seems likely that different proteins may respond to selection for greater thermostability in distinctly different ways. Despite the obvious complexities of the issue, we can expect widespread continued study of temperature adaptation at the molecular level, especially in proteins, because the results are not only of great biological interest but also of commercial and practical interest - both in the discovery of new, naturally occurring 'thermozymes' and in the design of new custom thermozymes for industrial purposes [58-60].

Acknowledgements

References

1. Russell AP, Holleman DS: **The thermal denaturation of DNA: average length and composition of denatured areas.** *Nucleic Acids Res* 1974, **1:**959–978.
2. Galtier N, Lobry JR: **Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes.** *J Mol Evol* 1997, **44:**632-636.
3. Hurst LD, Merchant AR: **High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes.** *Proc R Soc Lond B Biol Sci* 2001, **268:**493-497.
4. Forsdyke DR, Bell SJ: **Purine loading, stem-loops and Chargaff's second parity rule: a discussion of the application of elementary principles to early chemical observations.** *Appl Bioinformatics* 2004, **3:**3-8.
5. Muto A, Osawa S: **The guanine and cytosine content of genomic DNA and bacterial evolution.** *Proc Natl Acad Sci USA* 1987, **84:**166-169.
6. Singer GAC, Hickey DA: **Nucleotide bias causes a genomewide bias in the amino acid composition of proteins.** *Mol Biol Evol* 2000, **17:**1581-1588.
7. Sueoka N: **Wide intra-genomic G+C heterogeneity in human and chicken is mainly due to strand-symmetric directional mutation pressures: dGTP-oxidation and symmetric cytosine-deamination hypotheses.** *Gene* 2002, **300:**141-154.
8. Singer GAC, Hickey DA: **Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content.** *Gene* 2003, **317:**39-47.
9. Wang HC, Singer GAC, Hickey DA: **Mutational bias affects protein evolution in flowering plants.** *Mol Biol Evol* 2004, **21:**90-96.
10. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH: **Codon usage between genomes is constrained by genome-wide mutational processes.** *Proc Natl Acad Sci USA* 2004, **101:**3480-3485.
11. Forterre P: **A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein**. *Trends Genet* 2002, **18:**236-237.
12. Nakashima H, Fukuchi S, Nishikawa K: **Compositional changes in RNA, DNA and proteins for bacterial adaptation to higher and lower temperatures.** *J Biochem (Tokyo)* 2003, **133:**507-513.
13. Wang HC, Hickey DA: **Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes.** *Nucleic Acids Res* 2002, **30:**2501-2507.
14. Paz A, Mester D, Baca I, Nevo E, Korol A: **Adaptive role of increased frequency of polypurine tracts in mRNA sequences of thermophilic prokaryotes.** *Proc Natl Acad Sci USA* 2004, **101:**2951-2956.
15. Klein RJ, Misulovin Z, Eddy SR: **Noncoding RNA genes identified in AT-rich hyperthermophiles.** *Proc Natl Acad Sci USA* 2002, **99:**7542-7547.
16. Gutell RR, Cannone JJ, Shang Z, Du Y, Serra MJ: **A story: unpaired adenosine bases in ribosomal RNAs.** *J Mol Biol* 2000, **304:**335-354.
17. Lao PJ, Forsdyke DR: **Thermophilic bacteria strictly obey Szybalski's transcription direction rule and politely purine-load RNAs with both adenine and guanine.** *Genome Res* 2000, **10:**228-236.
18. Lambros RJ, Mortimer JR, Forsdyke DR: **Optimum growth temperature and the base composition of open reading frames in prokaryotes.** *Extremophiles* 2003, **7:**443-450.
19. Ikemura T: **Correlation between the abundance of** *Escherichia coli* **transfer RNAs and the occurrence of the respective codons in its protein genes.** *J Mol Biol* 1981, **146:**1-21.
20. Sharp PM, Stenico M, Peden JF, Lloyd AT: **Codon usage: mutational bias, translational selection, or both?** *Biochem Soc Trans* 1993, **21:**835-841.
21. Lobry JR, Chessel D: **Internal correspondence analysis of codon and amino-acid usage in thermophilic bacteria.** *J Appl Genet* 2003, **44:**235-261.
22. Rispe C, Delmotte F, van Ham RC, Moya A: **Mutational and selective pressures on codon and amino acid usage in** *Buchnera,* **endosymbiotic bacteria of aphids.** *Genome Res* 2004, **14:**44-53.
23. Kanaya S, Kinouchi M, Abe T, Kudo Y, Yamada Y, Nishi T, Mori H, Ikemura T: **Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the** *E. coli* **O157 genome.** *Gene* 2001, **276:**89-99.
24. Lynn DJ, Singer GAC, Hickey DA: **Synonymous codon usage is subject to selection in thermophilic bacteria.** *Nucleic Acids Res* 2002, **30:**4272-4277.
25. Jaenicke R, Böhm G: **The stability of proteins in extreme environments.** *Curr Opin Struct Biol* 1998, **8:**738-748.
26. Petsko GA: **Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat".** *Methods Enzymol* 2001, **334:**469-478.
27. Kreil DP, Ouzounis CA: **Identification of thermophilic species by the amino acid compositions deduced from their genomes.** *Nucleic Acids Res* 2001, **29:**1608-1615.
28. Tekaia F, Yeramian E, Dujon B: **Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis.** *Gene* 2002, **297:**51-60.
29. Farias ST, Bonato MC: **Preferred amino acids and thermostability.** *Genet Mol Res* 2003, **2:**383-393.
30. Friedman R, Drake JW, Hughes AL: **Genome-wide patterns of nucleotide substitution reveal stringent functional constraints on the protein sequences of thermophiles.** *Genetics* 2004, **167:**1507-1512.
31. Kawashima T, Amano N, Koike H, Makino S, Higuchi S, Kawashima-Ohya Y, Watanabe K, Yamazaki M, Kanehori K, Kawamoto T, *et al.*: **Archaeal adaptation to higher temperatures revealed by genomic sequence of** *Thermoplasma volcanium.* *Proc Natl Acad Sci USA* 2000, **97:**14257-14262.
32. Kumar S, Nussinov R: **How do thermophilic proteins deal with heat?** *Cell Mol Life Sci* 2001, **58:**1216-1233.
33. Suhre K, Claverie JM: **Genomic correlates of hyperthermostability, an update.** *J Biol Chem* 2003, **278:**17198-17202.
34. Thompson MJ, Eisenberg D: **Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability.** *J Mol Biol* 1999, **290:**595-604.
35. Zhang J. **Protein-length distributions for the three domains of life.** *Trends Genet* 2000, **16:**107-109.
36. Das R, Gerstein M: **The stability of thermophilic proteins: a study based on comprehensive genome comparison.** *Funct Integr Genomics* 2000, **1:**76-88.
37. Chakravarty S, Varadarajan R: **Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study.** *Biochemistry* 2002, **41:**8152-8161.
38. Alsop E, Silver M, Livesay DR.**Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis.** *Protein Eng* 2003, **16:**871-874.
39. Pack SP, Yoo YJ: **Protein thermostability: structure-based difference of amino acid between thermophilic and mesophilic proteins.** *J Biotechnol* 2004, **111:**269-277.
40. Kumar S, Nussinov R: **Fluctuations in ion pairs and their stabilities in proteins**. *Proteins* 2001, **43:**433-454.
41. Shockley KR, Ward DE, Chhabra SR, Conners SB, Montero CI, Kelly RM: **Heat shock response by the hyperthermophilic archaeon** *Pyrococcus furiosus.* *Appl Environ Microbiol* 2003, **69:**2365-2371.
42. Makarova KS, Wolf YI, Koonin EV: **Potential genomic determinants of hyperthermophily.** *Trends Genet* 2003, **19:**172-176.
43. Foster PG, Jermiin LS, Hickey DA. **Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria.** *J Mol Evol* 1997, **44:**282-288.
44. Knight RD, Freeland SJ, Landweber LF: **A simple model based on mutation and selection explains trends in codon and amino-**

acid usage and GC composition within and across genomes. *Genome Biol* 2001, **2:**research0010.1-0010.13.

45. Lobry JR: **Asymmetric substitution patterns in the two DNA strands of bacteria.** *Mol Biol Evol* 1996, **13:**660-665.

46. Lafay B, Lloyd AT, McLean MJ, Devine KM, Sharp PM, Wolfe KH: **Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases.** *Nucleic Acids Res* 1999, **27:**1642-1649.

47. Foster PG, Hickey DA: **Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions.** *J Mol Evol* 1999, **48:**284-290.

48. Phillips MJ, Delsuc F, Penny D: **Genome-scale phylogeny and the detection of systematic biases.** *Mol Biol Evol* 2004, **21:**1455-1458.

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

50. Roberts D: **Eukaryotic cells under extreme conditions**. In *Enigmatic Microorganisms and Life in Extreme Environments*. Edited by Seckbach J. Dordrecht: Kluwer; 1999, 163-173.

51. Tansey MR, Brock TD: **The upper temperature limit for eukaryotic organisms.** *Proc Natl Acad Sci USA* 1972, **69:**2426-2428.

52. Forterre P: **Thermoreduction, a hypothesis for the origin of prokaryotes**. *C R Acad Sci III* 1995, **318:**415-422.

53. Sprott GD: **Structures of archaebacterial membrane lipids.** *J Bioenerg Biomembr* 1992, **24:**555-566.

54. Portner HO: **Climate variations and the physiological basis of temperature dependent biogeography: systemic to molecular hierarchy of thermal tolerance in animals.** *Comp Biochem Physiol A Mol Integr Physiol* 2002, **132:**739-761.

55. Bernardi G: **Isochores and the evolutionary genomics of vertebrates.** *Gene* 2000, **241:**3-17.

56. Montoya-Burgos JI, Boursot P, Galtier N: **Recombination explains isochores in mammalian genomes.** *Trends Genet* 2003, **19:**128-130.

57. Meunier J, Duret L: **Recombination drives the evolution of GC-content in the human genome.** *Mol Biol Evol* 2004, **21:**984-990.

58. Vieille C, Zeikus GJ: **Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability.** *Microbiol Mol Biol Rev* 2001, **65:**1-43.

59. Haki GD, Rakshit SK: **Developments in industrially important thermostable enzymes: a review.** *Bioresour Technol* 2003, **89:**17-34.

60. Henne A, Bruggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R, *et al.*: **The genome sequence of the extreme thermophile *Thermus thermophilus.*** *Nat Biotechnol* 2004, **22:**547-553.

61. Huang SL, Wu LC, Liang HK, Pan KT, Horng JT, Ko MT: **PGTdb: a database providing growth temperatures of prokaryotes.** *Bioinformatics* 2004, **20:**276-278.

62. Korbel JO, Snel B, Huynen MA, Bork P: **SHOT: a web server for the construction of genome phylogenies.** *Trends Genet* 2002, **18:**158-162.