



## Short note

# DNA structure and architecture in the chromosome and plasmid of hyperthermophilic organisms, a theoretical approach

Héctor G. Vázquez<sup>1</sup>, Arturo Becerra<sup>1,\*</sup>

<sup>1</sup>Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, Delegación Coyoacán, 04510 CDMX, México.

\*abb@ciencias.unam.mx

### Abstract

Hyperthermophilic organisms have been recognized as an important element in the origin and early evolution of life on Earth, but also as a model for exobiological studies. Analyzing their molecular dynamics can help us understand this important lifestyle. In this study, using bioinformatic tools, the DNA composition in chromosomes and plasmids of hyperthermophilic organism were compared, and their structure, probable amino acid bias and DNA flexibility were analyzed. In some chromosomes and plasmids shows differential features of DNA flexibility and skews in mutation rate, which suggests that only some molecular elements show high values of variability, contrary to the proposal of the flexible genome theory.

Keywords: chromosome, plasmids, Quintana values, DNA twisting, codon usage, hyperthermophiles.

### Resumen

*Los organismos hipertermofílicos han sido reconocidos tanto como un elemento importante en el origen y la evolución temprana de la vida en la Tierra, como un modelo en estudios de exobiología. Analizar sus dinámicas moleculares nos puede ayudar a entender este importante estilo de vida. En este estudio, usando scripts y software, se comparó la composición del DNA en los cromosomas y plásmidos de organismos hipertermófilos, y se analizó su estructura, su probable sesgo en el uso de aminoácidos y la flexibilidad del DNA. En algunos cromosomas y plásmidos se presentaron características diferenciales de flexibilidad del DNA y sesgos en su tasa de mutación, lo cual sugiere que sólo algunos elementos moleculares muestran altos niveles de variabilidad, contrariamente a lo propuesto en la teoría de genoma flexible.*

*Palabras clave: cromosoma, plásmidos, valores de Quintana, torsión de DNA, uso de codones, hipertermófilos.*

## 1. Introduction

Studies on extremophilic microorganisms have been a breakthrough for the fields of biochemistry (Xu and Glansdorff, 2002; Connors *et al.*, 2006), biotechnology (Vieille and Zeikus, 2001; Guiral *et al.*, 2012) and early evolution, broadening our understanding of the limits of

life (Nisbet and Sleep, 2001; Allers and Mevarech, 2005).

Extremophiles are organisms that surpass the mesophilic limits of parameters such as temperature, pH and salinity, and can have biochemical dependencies on different organic and inorganic compounds like sulfur, nitrogen, methane or even ferrous oxides (respectively) (Stetter *et al.*, 1990). The respiration reactions of these organisms

increase our awareness of what we might find outside our planet (Trent, 2000). It is necessary to integrate a clear approach to the study of these organisms to recognize new trends and molecular signatures in different environments. One of the extremophilic lifestyles that has been heavily studied due to its impact in early evolution (Islas *et al.*, 2003) and its biochemical diversification is the group of hyperthermophilic prokaryotes from the Archaea and Bacteria domains (Stetter, 2006). A large number of studies involving hyperthermophilic prokaryotes are based on their multiple strategies for survival and their stability in high-temperature environments (Stetter *et al.*, 1990). There is no single molecule or metabolic pathway unique for this lifestyle (Atomi *et al.*, 2004), and finding a pattern or common properties in all the described species remains elusive (Trent, 2000; Allers and Mevarech, 2005). However, it is possible to associate some molecular traits with nucleotide and amino acid levels, beyond any phylogenetic groups with hyperthermophilic character. In this regard, Groussin and Gouy (2011) recognize two main characteristic processes in the hyperthermophilic lifestyle: *a*) the molecular evolutionary rate is skewed by the optimal growth temperature and tRNA-coding genes, and *b*) it is possible to correlate the optimal growth temperature with multiple components of coding and non-coding sequences. Also, Klipcan *et al.* (2006) correlated the optimal growth temperature with the proportion and type of aminoacyl-tRNA synthases (aatRNAs). Agarwal and Grover (2008) recognize a purine bias that modifies the amino acid frequency and codon usage. Therefore, it has not yet been possible to recognize common genes for all the hyperthermophilic genomes, but it is feasible to discover biases and proportions in amino acids and nucleotides that could define a differential composition in this lifestyle. As it has been proposed (Cordero and Polz, 2014), the hyperthermophilic genome can be studied in two main sections: 1) the core genome that is composed of all the housekeeping genes involved in basic metabolism, and 2) the flexible genome, which is shaped by genes related to habitat-specific properties, as well as interaction between viruses and predators. These genes have a variable presence in the entire genome, and are involved in horizontal gene transfer events, gene loss and high rates of gene turnovers.

With these pangenomic studies, it is possible to infer that the bacterial chromosomes are composed of genes from the core and flexible genomes, but the extrachromosomal materials, like the plasmids, are composed mainly of genes related with the flexible genome.

By studying the plasmid genome, it is possible to recognize the arrangement of the structure of the DNA and the variance of codon usage in the same organism. Berg and Kurland (2002) and Cordero and Polz (2014) proposed that the study of the flexible genome allows the identification of global or individual genes that can be used as the reflection of the prokaryotic community sharing the same environmental conditions. Also, according to Cooper

*et al.* (2010), secondary bioinformational elements (like secondary chromosomes, megaplasmids, or plasmids) show a decrease in codon usage diversity in the case of organisms living in the same environment, despite belonging to different phylogenetic groups.

Since codon usage have a differential appearance in the coding sequences, there are different ways to measure the structure and topology of the genome. One approach consists of recognizing dinucleotide interaction in the entire DNA sequence and correlating it with the DNA twist profile. As noted by Quintana *et al.* (1992), the DNA crystallography profiles correlate the DNA twist with the space and configuration among dimers. The **High twist profile** (H value) indicates the incidence of sequence of dinucleotides GC or GA, and elevated values of DNA twist and space among nucleotide dimers. In contrast, the **Low twist profile** (L value) is present where the sequence has a high frequency of CC, CT, TT, AA, AG and GG dimers. These configurations present the values of twist and separation between the dimers. The **Variable twist profile** (V value) is a combination of both configurations and is correlated with the incidence of pyrimidine and purine dimers, a conformation that is strongly susceptible to the influence of the environment.

This paper focuses primarily on the comparative genomics, with special emphasis on the evolution of the extremophilic genomes. For the codon usage analysis, we review Jaenicke *et al.* (1991), which implied that a decrease in non-charged amino acids allows discrimination between extremophile and mesophile proteomes. Also, Zeldovich *et al.* (2007) correlate the increase in the use of seven amino acids (IVYWREL) with a raise in the optimal growth temperature. Using both approaches—analysis in DNA twist profile and codon usage—we tried to identify a shared trait among the chromosomes and plasmids of some hyperthermophilic organisms. If found, we might recognize common characteristics for the hyperthermophilic lifestyle.

## 2. Methods

In order to classify and recognize the optimal temperature interval for the proposed archaeal and bacterial species, we use the National Center for Biotechnology and Information (NCBI) database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). These data are compiled in Table 1. A total of eight hyperthermophilic species with complete chromosome and plasmid sequences were obtained from the ftp site of NCBI. The eight species include three Crenarchaeota (*Sulfolobus islandicus* L.D.8.5, *Sulfolobus islandicus* Y.N.15.51, and *Thermofilum pendens* Hrk5), four Euryarchaeota (*Archaeoglobus profundus* DSM 5631, *Methanococcus maripaludis* C5, *Pyrococcus abyssi* and *Thermococcus barophilus* MP), and one species from Bacteria (*Aquifex aeolicus* VF5). A mesophilic organism (*Escherichia coli* O157 H7) was used as a negative control organism. We chose this strain because of their mesophilic

Table 1. Thermophilic and hyperthermophilic genomes of the Archaea and Bacteria domains used in this paper.

Subdivision	Species	Chromosome (c) and plasmid (p) size (Mb)	GC amount (%) for chromosomes (c) and plasmids (p)	Optimal growth temperature (°C)
Aquificales (Bacteria)	<i>Aquifex aeolicus</i> VF5	1.55 (c)	43.5 (c)	93
		0.039 (p)	36.4(p)	
Enterobacteriales (Bacteria)	<i>Escherichia coli</i> O157 H7	4.64 (c)	50.7 (c)	35
		0.09 (p)		
Crenarchaeota (Archaea)	<i>Sulfolobus islandicus</i> L.D.8.5.	2.72 (c)	35.3 (c)	85
		0.026 (p)	36.1 (p)	
	<i>Sulfolobus islandicus</i> Y.N.15.51	2.81 (c)	35.3(c)	85
		0.04 (p)	36.1 (p)	
Euryarchaeota (Archaea)	<i>Archaeoglobus profundus</i> DSM 5631	1.56 (c)	42 (c)	102
		0.0028 (p)	39.8 (p)	
	<i>Methanococcus maripaludis</i> C5	1.78 (c)	33.0 (c)	45
		0.008 (p)	27.2 (p)	
	<i>Pyrococcus abyssi</i>	1.77 (c)	44.7 (c)	102
		0.003 (p)	43.4 (p)	
<i>Thermococcus barophilus</i> MP	2.01 (c)	41.8 (c)	85	
	0.054 (p)	38.3 (p)		

Note. The data from the chromosomes (c) and plasmids (p) come from the NCBI database, and the characteristics and the optimal growth temperature from Horneck and Baumstark-Khan (2002).

lifestyle and the incidence of their unique plasmid structure and representation. The H, V and L mean values were evaluated on genomes and plasmids using the program codon.pl.

Additionally, using the UGENE 1.19 software (Okonechnikov *et al.*, 2012), we calculated high flexibility areas and tandem repeat sequences in both chromosomes and plasmids. The module that would recognize the high flexible sequences was applied using the default values, and the tandem repeat module was modified to recognize sequences greater than 20 nucleotides. The size of tandem repeats is based on previous reports (van der Oost *et al.*, 2014).

In this study, we selected only completely sequenced genomes of organisms with reported thermophilic or hyperthermophilic lifestyles (Stetter *et al.*, 1990; Horneck and Baumstark-Khan, 2002), separated into chromosome and plasmid. Also, the genome of *Methanococcus*

*maripaludis* was included because of their thermophilic tolerance, and it would be important to have a comparative example with a methanobacterial genome.

### 3. Results

#### 3.1. Genome sampling in the NCBI database

The information in Table 1 shows the available diversity of archaeal and bacterial hyperthermophilic species, with complete chromosome and plasmid sequences. We analyzed eight species: one from the bacterial domain, three from Crenarchaeota subdomain, and four from the Euryarchaeota subdomain. The comparison of the GC amount in the plasmids and chromosomes shows that this value in Crenarchaeota is more stable than in the other groups. In addition, two species with an increased genome size were

found. Crenarchaeota has a stable genomic structure, unlike the Euryarchaeota that has differences among the plasmid and chromosome GC values.

### 3.2. V, H and L comparative values

In Figure 1, the three main profiles of DNA twisting are integrated. The results cluster into the same area in which the H value increases in diverse chromosomes and plasmids, and the L value is positive and contains ~ 75 % of the sample. The only exceptions to this grouping are the *Thermofilum pendens* genome, the genome and plasmid of *Escherichia coli* O157 H7 and the *Methanococcus maripaludis* plasmid. Additionally, we identified a large difference in the V values

for the chromosome and plasmid of *Aquifex aeolicus* and the plasmid of *Archaeoglobus profundus*. This does not correspond with a phylogenetic signal, or a similar optimal growth temperature or even GC amount. The result was similar for *Aquifex* and *Archaeoglobus* plasmids that show overlapping and similar values of twisting.

### 3.3. Tandem repeat sequences and High flexibility regions

In order to recognize if all the plasmids show the same flexibility in their genome, or if they share a common aspect in this feature, we analyzed the incidence of particular regions and coupled it with tandem repeat regions (Allers and Mevarech, 2005; Norais *et al.*, 2013). The results are

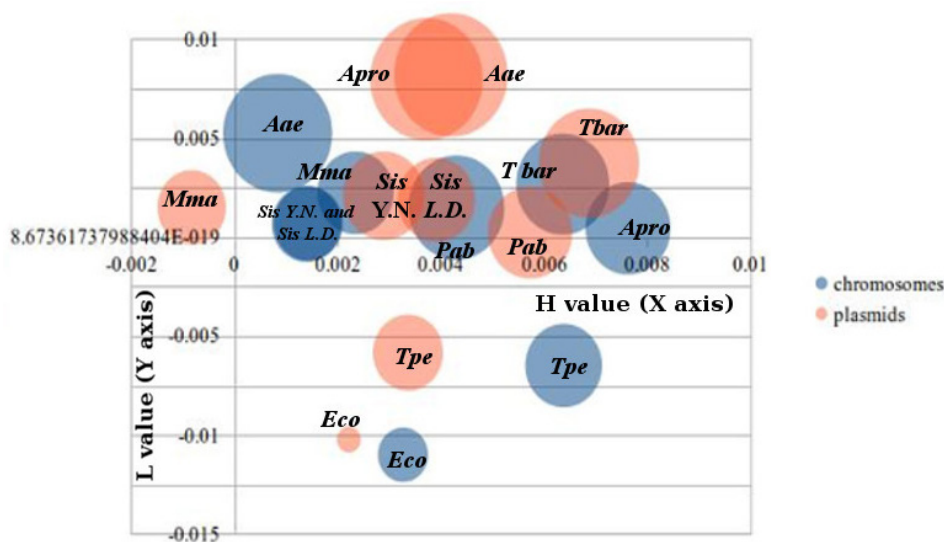


Figure 1. Graphical representation of the profile values for the chromosomes and plasmids of hyperthermophiles. The H and L values were used for abscises and ordinates, respectively, and the V value were used as an absolute number for bubble size representation. All values in %. The name of the species are pointed as follows: Aae = *Aquifex aeolicus* VF5; Sis = *Sulfolobus islandicus* (L.D. = L.D.8.5.; Y.N. = Y.N.15.51); Tpe = *Thermofilum pendens* Hrk5; Apro = *Archaeoglobus profundus* DSM 5631; Mma = *Methanococcus maripaludis* C5; Pab = *Pyrococcus abyssi* and Tbar = *Thermococcus barophilus* MP; Eco = *Escherichia coli* O157 H7. Obtained values for chromosomes are in blue, and for plasmids in red.

Table 2. Tandem repeat sequences identified in genomic hyperthermophiles.

Species and bioinformational molecule	Tandem repeat sequences	
	Length (number of nucleotides) and number of sequences in parentheses.	Example
<i>Aquifex aeolicus</i> VF5 (chromosome)	67 (2)	- 5' GATTGGGAAATTTTTTTTAAAT 3' -
<i>Methanococcus maripaludis</i> (chromosome)	41 (7), 315 (1), 644 (1), and 381 (1)	- 5' GCTGTCCTGTTAAGCATTCTT 3' -
<i>Sulfolobus islandicus</i> Y.N.15.51 (chromosome)	54 (2) and 99 – 105 (4)	- 5' GTCAAACGCCATTGCATAT 3' -
<i>Sulfolobus islandicus</i> L.D.8.5. (chromosome)	60 (2)	- 5' GAAGTTTTAGTTTCTTTTCG 3' -

Note. Only sequences and regions with the default values of UGENE are recognized.

Table 3. High flexibility regions identified in genomic hyperthermophiles.

Species and bioinformational molecule	High flexibility sequences	
	Length (number of nucleotides) and number of sequences in parentheses.	Example
<i>Methanococcus maripaludis</i> (chromosome)	103 – 106 (5); 141 (2) 120 (1) and 153 (1).	- 5' GTAATATTAATTTTAATA 3' -
<i>Methanococcus maripaludis</i> (plasmid)	130 (1).	- 5' GATATTTTTTTATATATTT 3' -
<i>Sulfolobus islandicus</i> Y.N.15.51 (chromosome)	101 – 110 (7); 118 (1). 128 – 133 (2). and 174 (1).	- 5' GATATATTTGGTGGTTA 3' -
<i>Sulfolobus islandicus</i> L.D.8.5 (chromosome)	32 (1); 60 (2); 101 – 107 (3). 116 (1); 125 (2) 174 (1).	- 5' GTAAATATATGCATATAA 3'

Note. Only sequences and regions with the default values of UGENE are recognized.

shown in Table 2 and 3. Contrary to previous models, high flexibility values for plasmids were not recognized. Only the plasmid of *M. maripaludis* has high flexibility and shares a similar L value with the chromosome. The structure of the chromosome allows both high flexibility regions and tandem repeat regions. The chromosome of both *Sulfolobus* species shows both elements. This signal is not shared by any of the analyzed hyperthermophilic genomes.

#### 3.4. Codon usage and amino acid values

For the correlating values, we notice that the highest values occur among plasmids (Figure 2), with  $R^2 > 0.7$  for leucine and isoleucine, whereas the highest value for chromosomes (Figure 3) is shown by glutamic acid ( $R^2 \approx 0.7$ ).

## 4. Discussion

### 4.1. Genome sampling and further samples

Although the sample used is relatively small, important conclusions can be drawn. This work provides an approach as to how the hyperthermophilic Bacteria and Archaea can be arranged for DNA twisting in chromosomes and plasmids. One case that needs further study is the arrangement of Sulfolobales, where both sets of analyzed chromosomes have an identical DNA twist value (for H and

L values), and a different value for their plasmids (implied in the H value). The Sulfolobales order still needs further study, including different species like *S. acidocaldarius* to confirm their structure and variation as a phylogenetic trend that could complement the previously reported high flexibility genomes (Zillig *et al.*, 1996; Farkas *et al.*, 2011).

Increasing the amount of complete hyperthermophilic genomes, especially bacterial, will provide a better understanding of the flexible genome perspective.

### 4.2. H, L and V values, and their correlation with the hyperthermophilic lifestyle

Although not all values are correlated directly with the hyperthermophilic lifestyle, we can recognize similar V and L values; this applies to 75 % of the sample. This proportion occurs in both chromosomes and plasmids, and may imply that the amount related to Low twist profile (L) arrangement and Variable twist profile (V) in the archaeal hyperthermophilic genome is a main trend.

The eccentric position of the *T. pendens* genome could be explained by the high incidence of events of gene loss (Anderson *et al.*, 2008), and the effect of a recent split transfer genes involving informational genes (Chan *et al.*, 2011). *M. maripaludis* shows a negative H value, and it is possible that the DNA twist and structure are associated with its thermotolerant and not hyperthermophilic lifestyle. Furthermore, it has been reported that the Methanobacteriales have a high incidence of gene conversion (Hildenbrand

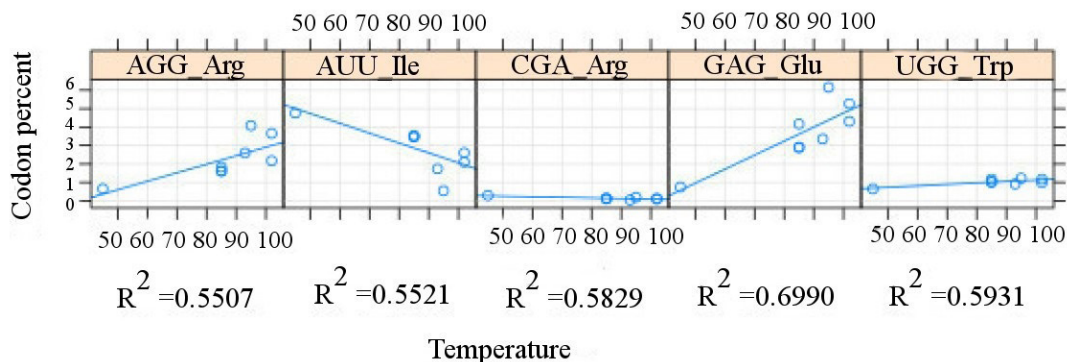


Figure 2. Correlation between the codon usage in chromosomes and the optimal growth temperature (OGT). The relation between the reported optimal growth temperature (abscises) versus the percent value (ordinates), as evaluated from the overall diversity of the genetic code, is presented. The  $R^2$  value of each resulting codon is shown below the corresponding area.

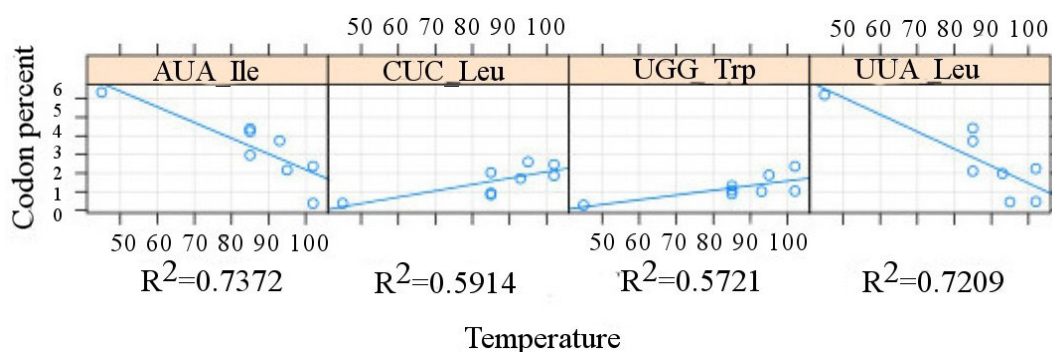


Figure 3. Correlation between the codon usage in plasmids and the optimal growth temperature (OGT). The relation between the reported optimal growth temperature (abscises) versus the percent value (ordinates), as evaluated from the overall diversity of the genetic code, is presented. The  $R^2$  value of each resulting codon is shown below the corresponding area.

*et al.*, 2011) that could modify their base composition and gene arrangements. After integrating *Escherichia coli* O157 H7 as a mesophilic control, its proportion and incidence, graphed in a different area of Figure 1, provides a good comparison with mesophilic values.

The nearness of the V and H values of plasmids from *Archaeoglobus profundus* and *Aquifex aeolicus* could be evidence of the high horizontal gene transfer events between them (van Wolferen *et al.*, 2013). However, it is necessary to develop further pangenomic analysis and comparative studies of these sequences.

#### 4.3. Tandem repeats and High flexibility sequences

The highest incidence of tandem repeats and high flexibility sequences in chromosomes allowed us to identify punctual regions involved in recombination and increased flexibility and possibly "hotspots" of mutation and recombination.

The finding of only one sequence with repeats in the plasmid of *M. maripaludis* suggests that plasmids from hyperthermophilic species need additional analyses. This result contrasts with the general idea in which tandem repeats and high flexibility sequences are correlated with

recombination events (Johnson *et al.*, 2013) and with the occurrence of DNA repair mechanisms (Cai *et al.*, 2009) that shift the structure of the plasmid DNA. Moreover, this suggests that plasmids in hyperthermophilic organisms might be regulated by different mechanisms in their chromosomes.

#### 4.4. Codon usage and amino acid proportion

By comparing the codon usage for the seven proposed amino acids that correlate with thermotolerance, we have a different pattern than previously published by Zeldovich (Zeldovich *et al.*, 2007). Although it is impossible to associate the increase of certain codon amino acids to the role of thermal stabilizers, this brings us an additional point of view about the genome structure response to the extremophilic lifestyle.

The proposed analysis allowed the recognition of different patterns in the same amino acid. An example of this differential pattern is the decreasing isoleucine usage, which is negatively correlated for chromosome and plasmid for all the analyzed hyperthermophiles, although with different codons. On the other hand, leucine is positively correlated with OGT (codon CUC) and negatively in plasmids (with



codon UUA). This disagrees with Zeldovich *et al.* (2007), and consequently a more detailed analysis is required. That would include other extremophilic groups.

Furthermore, the increase in glutamic acid is significant in chromosomes along with an increase in OGT, while in the plasmid the UGG codon (tryptophan) increases with OGT. It has been proposed that glutamic acid is relevant for hyperthermophilic organisms because of its charge, thus causing a difference in side chain entropy, and helping in protein folding under extreme conditions (Greaves and Warwicker, 2007). However, tryptophan is an amino acid whose amount decreases into thermostable proteins (de Champdoré *et al.*, 2007), and shows a small increase in plasmid sequences.

With this, we infer that proteins coded in the chromosomes show different performances from those coded in the plasmid. This bias in the plasmid could be explained by their accessory role in the metabolism of hyperthermophilic organisms.

## 5. Conclusions

The study of the dynamic DNA structure of archaeal and bacterial hyperthermophiles allows further understanding of the biology of this particular lifestyle. We determined that the H, V and L values are similar for all the analyzed organisms.

There is not satisfactory correlation on the changing trend of codon usage in chromosomes and plasmids with skews in the coding sequence.

## Acknowledgements

This paper constitutes a partial fulfillment of the Programa de Posgrado en Ciencias Biológicas of the Universidad Nacional Autónoma de México (UNAM). H.V. acknowledges the scholarship and financial support provided by the National Council of Science and Technology (CONACyT), and UNAM. We thank the support and ideas from Dr. Pedro Miramontes. Also, we are thankful to Dr. Germinal Cocho for support in the knowledge and understanding of DNA twist and flexibility concepts, and Dr. Luis Delaye for the use and support of his script and programming abilities.

We thank the reviewers for all the support and all the constructive observations that made possible the adequate comprehension and final development of this article.

## References

Agarwal, S., Grover, A., 2008, Nucleotide composition and amino acid usage in AT-rich hyperthermophilic species: *Open Bioinformatics Journal*, 2, 11-19.

- Allers, T., Mevarech, M., 2005, Archaeal genetics - the third way: *Nature Reviews Genetics*, 6(1), 58-73.
- Anderson, I., Rodriguez, J., Susanti, D., Porat, I., Reich, C., Ulrich, L., Elkins, J., Mavromatis, K., Lykidis, A., Kim, E., Thompson, L., Nolan, M., Land, M., Copeland, A., Lapidus, A., Lucas, S., Detter, C., Zhulin, I., Olsen, G., Whitman, W., Mukhopadhyay, B., Bristow, J., Kyrpides, N., 2008, Genome sequence of *Thermofilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction: *Journal of Bacteriology*, 190, 2957-2965.
- Atomi, H., Matsumi, R., Imanaka, T., 2004, Reverse gyrase is not a prerequisite for hyperthermophilic life: *Journal of Bacteriology*, 186, 4829-4833.
- Berg, O., Kurland, C., 2002, Evolution of microbial genomes: sequence acquisition and loss: *Molecular Biology and Evolution*, 19, 2265-2276.
- Cai, Y., Patel, D.J., Geacintov, N.E., Broyde, S., 2009, Differential nucleotide excision repair susceptibility of bulky DNA adducts in different sequence contexts: hierarchies of recognition signals: *Journal of Molecular Biology*, 385, 30-44.
- Chan, P.P., Cozen, A.E., Lowe, T.M., 2011, Discovery of permuted and recently split transfer RNAs in Archaea: *Genome Biology*, 12, R38.
- Connors, S., Mongodin, E., Johnson, M., Montero, C., Nelson, K., Kelly, R., 2006, Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species: *FEMS Microbiology Reviews*, 30, 872-905.
- Cooper, V., Vohr, S., Wrocklage, S., Hatcher, P., 2010, Why genes evolve faster on secondary chromosomes in bacteria: *PLoS computational biology*, 6, e1000732.
- Cordero, O., Polz, M., 2014, Explaining microbial genomic diversity in light of evolutionary ecology: *Nature Reviews Microbiology*, 12, 263-273.
- de Champdoré, M., Staiano, M., Rossi, M., D'Auria, S., 2007, Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest: *Journal of the Royal Society, Interface / the Royal Society*, 4, 183-191.
- Farkas, J., Chung, D., DeBarry, M., Adams, M., Westpheling, J., 2011, Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon *Pyrococcus furiosus* needed for construction of a stable replicating shuttle vector: *Applied and Environmental Microbiology*, 77, 6343-6349.
- Greaves, R.B., Warwicker, J., 2007, Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles: *BMC Structural Biology*, 7, e18.
- Groussin, M., Gouy, M., 2011, Adaptation to Environmental Temperature is a Major Determinant of Molecular Evolutionary Rates in Archaea: *Molecular Biology and Evolution*, 28, 1-42.
- Guiral, M., Prunetti, L., Aussignargues, C., Ciaccafava, A., Infossi, P., Ilbert, M., Lojou, E., Giudici-Orticoni, M., 2012, The hyperthermophilic bacterium *Aquifex aeolicus*: from respiratory pathways to extremely resistant enzymes and biotechnological applications: *Advances in Microbial Physiology*, 61, 125-194.
- Hildenbrand, C., Stock, T., Lange, C., Rother, M., Soppa, J., 2011, Genome copy numbers and gene conversion in methanogenic archaea: *Journal of Bacteriology*, 193, 734-743.
- Homeck, G., Baumstark-Khan, C., 2002, *Astrobiology: Berlin, Heidelberg, Springer*, 411 p.
- Islas, S., Velasco, A., Becerra, A., Delaye, L., Lazcano, A., 2003, Hyperthermophily and the origin and earliest evolution of life: *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, 6, 87-94.
- Jaenicke, R., 1991, Protein stability and molecular adaptation to extreme conditions: *European Journal of Biochemistry / FEBS*, 202, 715-728.
- Johnson, S., Chen, Y.-J., Phillips, R., 2013, Poly(dA:dT)-rich DNAs are highly flexible in the context of DNA looping: *PLoS one*, 8, e75799.
- Klipcan, L., Safro, I., Temkin, B., Safro, M., 2006, Optimal growth temperature of prokaryotes correlates with class II amino acid composition: *FEBS letters*, 580, 1672-1676.
- Nisbet, E., Sleep, N., 2001, The habitat and nature of early life: *Nature*, 409, 1083-1091.

- Norais, C., Moisan, A., Gaspin, C., Clouet-d'Orval, B., 2013, Diversity of CRISPR systems in the euryarchaeal Pyrococcales: RNA biology, 10, 659-70.
- Okonechnikov, K., Golosova, O., Fursov, M., 2012, Unipro UGENE: a unified bioinformatics toolkit: Bioinformatics, 28, 1166-1167.
- Quintana, J., Grzeskowiak, K., Yanagi, K., Dickerson, R., 1992, Structure of a B-DNA decamer with a central T-A step: C-G-A-T-T-A-A-T-C-G: Journal of Molecular Evolution, 225, 379-395.
- Stetter, K., 2006, Hyperthermophiles in the history of life: Philosophical transactions of the Royal Society of London, Series B, Biological sciences, 361, 1474.
- Stetter, K., Fiala, G., Huber, G., Huber, R., Seegerer, A., 1990, Hyperthermophilic microorganisms: FEMS Microbiology Reviews, 75, 117-124.
- Trent, J., 2000, Extremophiles in astrobiology: per Ardua ad Astra: Gravitational and space biology bulletin publication of the American Society for Gravitational and Space Biology, 13, 5-11.
- Van der Oost, J., Westra, E., Jackson, R., Wiedenheft, B., 2014, Unravelling the structural and mechanistic basis of CRISPR-Cas systems: Nature reviews Microbiology, 12(7), 479-492.
- Van Wolferen, M., Ajon, M., Driessen, A. J. M., Albers, S.-V., 2013, How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions: Extremophiles, 17, 545-563.
- Vieille, C., Zeikus, G., 2001, Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability: Microbiology and Molecular Biology reviews, 65, 1-43.
- Xu, Y., Glansdorff, N., 2002, Was our ancestor a hyperthermophilic procaryote? Comparative biochemistry and physiology. Part A: Molecular & Integrative Physiology, 133, 677-688.
- Zeldovich, K., Berezovsky, I., Shakhnovich, E., 2007, Protein and DNA sequence determinants of thermophilic adaptation: PLoS computational biology, 3, e5.
- Zillig, W., Prangishvili, D., Schleper, C., Elferink, M., Holz, I., Albers, S., Janekovic, D., Götz, D., 1996, Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea: FEMS microbiology reviews, 18, 225-36.
- Manuscript received: November 11, 2014  
Corrected manuscript received: March 10, 2015  
Manuscript accepted: March 24, 2015