

GENOMIC AND PROTEOMIC SIGNATURES OF RADIATION AND THERMOPHILIC ADAPTATION IN THE *DEINOCOCCUS-THERMUS* GENOMES

RACHANA BANERJEE*, AYAN ROY, SUBHASIS MUKHOPADHYAY

Department of Biophysics, Molecular Biology and Bioinformatics and Distributed Information Centre for Bioinformatics, University of Calcutta, 92, A.P. C. Road, Kolkata 700009. Email: rachanabanerjee23@yahoo.co.in

Received: 16 Jan 2014, Revised and Accepted: 28 Feb 2014

ABSTRACT

Objective: The bacterial phylum *Deinococcus-Thermus* includes two groups of bacterial genomes i.e., radiation-resistant mesophilic *Deinococcales* and radiation-sensitive hyperthermophilic *Thermales*. Our endeavor is to address the factors influencing the niche-specific molecular evolution of this immensely interesting bacterial phylum.

Methods: An extensive comparative analysis of codon and amino acid usage patterns, aided by multivariate statistical techniques, was carried out in our chore. Evolutionary analysis was also executed to investigate the riddles of adaptive strategy opted by this fascinating pool of microbes.

Results: Significant differences in synonymous codon usage bias and amino acid usage patterns have been reported between the hyperthermophilic, radiation-sensitive *Thermales* genomes and mesophilic, radiation-resistant *Deinococcales* genomes. The amino acid usage patterns of radiation-resistant, mesophilic organisms under study are primarily influenced by their mode of adaptation to radiation. Core proteins from the group show signatures of both purifying as well as positive selection. Housekeeping genes tend to be under the influence of strong purifying selection, whereas, proteins mainly associated with defense mechanisms and secretion are inferred to be under the force of positive adaptation.

Conclusion: G+C compositional bias has been revealed to be a crucial determinant in shaping the codon usage signature. Investigation at proteomic level reflects the fact that genomes of the phylum *Deinococcus-Thermus* have distinct amino-acid compositional feature, physicochemical and structural trait and varying degrees of stability of their core proteome, according to their adaptation towards radiation and variable temperature stresses. The evolutionary analysis hints to the adaptive strategy that may be employed by these groups of microbes to survive against extremities of temperature and radiation.

Keywords: Radiation-resistance, Hyperthermophilic, Relative synonymous codon usage, Amino-acid usage, Physico-chemical properties, Structural-traits.

INTRODUCTION

Fluctuating environmental condition is the most important driving force for the microbial evolution. Microbes are best known for their adaptability to varying lifestyle along with an extreme variation of environmental condition. The genomic architecture of a microbe may bear the signatures, featuring the specific niche in which they are adapted [1-6]. Unveiling these niche-specific signature patterns both at the genomic as well as proteomic level, is the major field of interest for the present-day researchers.

The bacterial phylum *Deinococcus-Thermus* comprises of bacteria that are highly resistant to environmental hazards [7]. *Deinococcus spp.* and *Thermus spp.* are two groups present in the *Deinococcus-Thermus* phylum. Phylogenetic trees, based on ribosomal RNAs and several other conserved proteins, support the common origin of these two groups [8-11]. The *Thermus* genus consists of thermophilic species, whereas the *Deinococcus* genus comprises of mesophilic species with extreme resistance to γ -irradiation and desiccation [12-14]. The order *Deinococcales* includes two genera i.e., *Deinococcus* and *Truepera*. Radiation- and desiccation-resistant bacteria have considerably lower protein oxidation levels than the radiation-sensitive bacteria, but have similar yields of DNA double-strand breaks. These findings pointed towards the protection of proteins against oxidative damage as a new paradigm of radiation toxicity and defied the concept of DNA as the primary target of radiation toxicity [15-19]. *Deinococcus radiodurans* is unique among all the other radiation-resistant *Deinococcus* species from the perspective of its ability to overcome oxidative stress that affects all cellular macromolecules [20-22]. Most of the *Deinococcus* species are mesophilic in nature with growth temperature ranges between 25°C to 35°C. But a few of them have slightly higher growth temperature i.e., between 35°C to 45°C [23-25]. *Truepera radiovictrix* is a thermophilic bacterium with radiation resistance property, which has been recently included in the group *Deinococcus* [26]. The order *Thermales* includes several genera resistant to heat but sensitive to radiation (*Marinithermus*, *Meiothermus*, *Oceanithermus*, and *Thermus*). Among these thermophilic genera,

Meiothermus is distinguished from the other by their slightly lower temperature optima i.e., 60°C [27] and the genera *Marinithermus* and *Oceanithermus* comprise of thermophilic species [28, 29]. But the bacterial species present within the genus *Thermus* are hyperthermophilic in nature, with optimum growth temperature of about 70°C [30-33]. Extensive research has been carried out in the past using the members of the phylum *Deinococcus-Thermus*, exploring their mechanism of thermophilic adaptations and radiation resistance. Comparative genomic approach has been used by researchers to detect the "thermophilic determinants" in the proteomes of thermophiles [10, 34-36]. Various propositions have also been offered to explain the radiation resistance in bacteria. Some of these propositions explained the role of specialized genetic systems for DNA repair and stress response [24]. Recently, alternative possibilities have also been proposed, which consider post-irradiation adjustment of metabolism of *D. radiodurans* for avoiding the production of reactive oxygen species (ROS), which is attained by decreasing the number of reactions involving oxygen and high intracellular manganese concentrations of *Deinococcus sp.* It helps to scavenge the ROS generated during irradiation and post-irradiation recovery [15, 16, 18, 19]. Omelchenko *et al.*, [10] had undertaken an extensive comparative analysis using two genomes i.e., *Thermus thermophilus* and *Deinococcus radiodurans*. Using the two genomes, they have characterized the radiation and desiccation resistance from an experimental view point and assessed the contribution of different evolutionary processes for the discrete adaptation of *Thermus thermophilus* and *Deinococcus radiodurans*. They have mentioned that divergence of genomes from their common ancestor can take place via various processes. For example, differential gene loss and gain, acquisition of genes via horizontal gene transfer (HGT) followed by loss of the ancestral orthologous gene, lineage-specific expansion of paralogous gene families by duplication and/or acquisition of paralogs via HGT, modification of amino acid composition that could affect protein stability etc. Other differences between thermophilic and mesophilic ones can manifest due to differences in their structural properties, such as different amino acid compositions, loop lengths, number of salt bridges,

strength of hydrophobic interaction, number of disulfide bonds etc. [37-42]. Identifying the unique feature of the gene repertoires of *Thermus thermophilus* and *Deinococcus radiodurans* and analyzing the proteins predicted to be involved in DNA repair and stress response functions i.e., particularly relevant to the adaptive evolution of resistance phenotypes, they have concluded that *Thermus thermophilus* and *Deinococcus radiodurans* have evolved from a common ancestor, the mechanisms of their resistance appear to be largely independent [10]. Till date no comparative analysis has been attempted using the complete phylum *Deinococcus-Thermus*. Due to fast sequencing technique, abundant complete and annotated genome sequences are available nowadays for the phylum *Deinococcus-Thermus*. In the present study, we have carried out a comparative analysis for identifying the molecular signatures exhibited by the members of the phylum *Deinococcus-Thermus* according to their capacities of various extents to resist radiation as well as thermal stress. After getting separated from the common ancestor, the two orders *Deinococcales* and *Thermales* have separately followed their own course of evolution and both of them have experienced a step-wise diversification according to the stressed condition they were up against. We have also pointed out that molecular pattern observed by us among these organisms is not simply the outcome of a continuous genetic drift. But, adequate adaptive selection forces and synonymous mutational pressure acting on their genome are equally influencing them in their gradual shift towards attaining the optimum level of adaptation in their own niche area.

MATERIALS AND METHODS

Sequence retrieval

Nucleotide and protein coding sequences of fully finished genomes of *Deinococcus-Thermus* (till September 2013, 19 completely annotated *Deinococcus-Thermus* genomes were available) were retrieved from the NCBI GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Some more *Deinococcus* genomes are also present in publicly available databases which appear to be permanent or draft, and as a result are not included in our study. For control, the protein coding sequences of *Aquifex aeolicus* (NC_000918.1), *Candidatus Nitrospira defluvii* (NC_014355.1), *Thermomicrobium roseum DSM 5159* (NC_011959.1), *Salinispora arenicola CNS-205* (NC_009953.1) and *Rubrobacter xylanophilus* (NC_008148.1) were also retrieved from GenBank. For our analysis, we have considered only those annotated ORFs, which encode proteins containing more than 100 amino acids. For the ease of description, we have used the term 'RR' for radiation-resistance organisms and 'RS' for radiation-sensitive organisms. Also, hyperthermophilic organisms, thermophilic organisms, slightly thermophilic organisms, comparatively low temperature adapted organisms among the *Thermus* group and mesophilic organisms have been termed as HT, T, ST, LT and M respectively. Descriptions about all 19 genomes from the phylum *Deinococcus-Thermus*, including the abbreviations used for them in the present study have been listed in Table 1.

Table 1: Detailed features of 19 genomes of the phylum *Deinococcus-Thermus* used in the study

Order	Name	Abbreviation	Mode of adaptation	OGT	GC-content	Size	Accession no. (Ref_Seq)	No. of proteins coded
<i>Deinococcales</i>	<i>Deinococcus deserti</i>	ddr	RR_M	30° C.	62.98	3.86	NC_012526.1	3459
<i>Deinococcales</i>	<i>Deinococcus proteolyticus</i>	dpt	RR_M	30° C.	65.67	2.89	NC_015161.1	2789
<i>Deinococcales</i>	<i>Deinococcus radiodurans</i>	dra	RR_M	30° C.	66.61	3.28	NC_001263.1	3167
<i>Deinococcales</i>	<i>Deinococcus peraridilitoris</i>	dpd	RR_M	30° C.	63.71	4.51	NC_019793.1	4183
<i>Deinococcales</i>	<i>Deinococcus gobiensis</i>	dgo	RR_M	30° C.	69.15	4.41	NC_017790.1	4340
<i>Deinococcales</i>	<i>Deinococcus maricopenis</i>	dmr	RR_ST	47.5°C	69.8	3.5	NC_014958.1	3264
<i>Deinococcales</i>	<i>Deinococcus geothermalis</i>	dge	RR_ST	47.5°C	66.44	3.25	NC_008025.1	3054
<i>Deinococcales</i>	<i>Truepera radiovictrix</i>	tra	RR_T	55°C	68.1	3.26	NC_014221.1	2945
<i>Thermales</i>	<i>Meiothermus ruber</i>	mrb	RS_LT	60°C	63.4	3.1	NC_013946.1	3014
<i>Thermales</i>	<i>Meiothermus silvanus</i>	msv	RS_LT	60°C	62.75	3.72	NC_014212.1	3505
<i>Thermales</i>	<i>Marinithermus hydrothermalis</i>	mhd	RS_T	68°C	68.1	2.27	NC_015387.1	2205
<i>Thermales</i>	<i>Oceanithermus profundus</i>	opr	RS_T	68°C	69.79	2.44	NC_014761.1	2373
<i>Thermales</i>	<i>Thermus sp. CCB_US3_UF1</i>	thc	RS_HT	75°C	68.57	2.26	NC_017278.1	2279
<i>Thermales</i>	<i>Thermus scotoductus</i>	tsc	RS_HT	75°C	64.99	2.36	CP001962.1	2458
<i>Thermales</i>	<i>Thermus thermophilus HB27</i>	tth	RS_HT	75°C	69.38	2.13	NC_005835.1	2210
<i>Thermales</i>	<i>Thermus thermophilus HB8</i>	tjt	RS_HT	75°C	69.49	2.12	NC_006461.1	2235
<i>Thermales</i>	<i>Thermus thermophilus JL-18</i>	ttl	RS_HT	75°C	68.98	2.31	NC_017587.1	2402
<i>Thermales</i>	<i>Thermus thermophilus SG0.5JP17</i>	tts	RS_HT	75°C	68.61	2.3	NC_017272.1	2339
<i>Thermales</i>	<i>Thermus oshimai</i>	tos	RS_HT	75°C	68.6	2.4	NC_019386.1	2111

** Descriptions of the mode of adaptations are as follows. RR_M=radiation-resistant, mesophilic, RR_ST=radiation-resistant, slightly thermophilic, RR_T= radiation-resistant, thermophilic, RS_LT= radiation-sensitive, comparatively low temperature adapted among the members of *Thermus* group, RS_T= radiation-sensitive, thermophilic, RS_HT= radiation-sensitive, hyperthermophilic. Optimum growth temperature has been abbreviated as OGT.

Multivariate analyses on synonymous codon and amino acid usage

Correspondence analysis (COA) is an ordination technique that identifies the key trends in the variation of the data and then in accordance with those trends it distributes the genes along continuous axes. It does not make any assumption about the data falling into discrete clusters and, therefore, represents continuous variation accurately [43-46]. Correspondence analysis on relative synonymous codon usage (RSCU) and amino acid usage of genes and proteins had been carried out on individual genomes using the program CODONW 1.4.2, in order to identify any significant variation in the usage of codons or amino acids among the genomes of the phylum *Deinococcus-Thermus* under study [47]. The pictorial illustration of the amino acid usage of all the genomes has been represented using a Heat map. Every column of the Heat map represents a color gradient from red to green, which stands for increasing values of frequency for a single amino acid. Correlation

analysis (Spearman's rank correlation and Pearson's correlation) and variance analysis was performed (at level of significance $P < 0.01$ or $P < 0.05$) using SPSS software package version 15.0.

Cluster analysis on amino acid usage

To find out the variation in amino acid usage between radiation-sensitive (RS) and radiation-resistant (RR) *Deinococcus-Thermus* genomes and also taking into account their thermophilic adaptability, a cluster analysis on standardized amino acid usage was carried out using STATISTICA (version 6.0, published by Statsoft Inc., USA) for all 19 *Deinococcus-Thermus* genomes (Table 1). To standardize the amino acid usage, we have used one mesophilic actinobacterial species viz., *Salinispora arenicola* having GC-content similar to those of the hyperthermophilic radiation-sensitive order *Thermales*, a hyperthermophilic bacterial species viz., *Candidatus Nitrospira defluvii* with GC-content less than the

hyperthermophilic radiation-sensitive order Thermales and one green non-sulphur bacteria i.e., Thermomicrobium roseum with GC-content similar to the radiation-resistant mesophilic order Deinococcales. We have considered the phylum hyperthermophilic bacteria, green non-sulphur bacteria and Actinobacteria as the reference for comparison, because they are evolutionarily close to the phylum Deinococcus-Thermus.

Determination of orthologs

Core genes among the 19 strains of *Deinococcus-Thermus* were screened using the CMG-Biotools workbench [48]. Orthologous proteins were selected on the basis of identity $\geq 50\%$ over at least 50% of the alignable region and E-value threshold of $1e-10$. A BLAST matrix was also generated that is included in Supplementary Figure 1. BLAST matrix is a simple comparative representation of proteins that are shared by the sets of genomes under scrutiny. The estimation of core gene set results to a count of 417 proteins that can be termed as the set of core proteomes making up the final orthologous cluster.

Estimation of synonymous and non-synonymous substitution patterns in orthologous sequences

Selection pressure acting on a protein-coding gene can be inferred from the value of the dN/dS ratio i.e., the ratio of the number of non-synonymous substitutions per non-synonymous site (dN) to the number of synonymous substitutions per synonymous site (dS). Higher proportion of non-synonymous over synonymous substitutions per site i.e., $dN/dS > 1$, point towards positive selection. We have calculated the dN and dS values using the yn00 program of the software PAML (for Phylogenetic Analysis by Maximum Likelihood) for all the 417 orthologs from three groups. The first group includes radiation-resistant order Deinococcales. The second group includes radiation-sensitive order Thermales, adapted to comparatively low temperature. While the third group comprises of radiation-sensitive order Thermales adapted to thermophilic as well as hyperthermophilic condition. The ad hoc method implemented in the program accounts for the transition/transversion rate bias and codon usage bias, and is an approximation to the Maximum Likelihood method accounting for the transition/transversion rate ratio [49-51].

Calculation of codon/amino acid usage indices and estimation of secondary structure of proteins

To find out the factors influencing codon and amino acid usage, we have calculated several indices like relative synonymous codon usage (RSCU) [52], G+C-content at third codon positions, G+C-content at the first and second codon positions i.e., (G+C)3% and (G+C)12% respectively, aromaticity and average hydrophobicity (Gravy score) (Kyte and Doolittle 1982) of protein coding sequences. The isoelectric point (pI) and instability index [53] of each protein were calculated using the ExPASy proteomics server [54]. Using the software PREDATOR [55], secondary structures (viz., helices, sheets, and coils) of the orthologs were computed and the variations in their percentage were also calculated.

COG classification

The predicted COGs annotations for each protein of all the 19 organisms have been done with the help of "The integrated microbial genomes and metagenomes (IMG/M) system" (<http://img.jgi.doe.gov/m/>), which provides a support for comparative analysis of metagenomes in a comprehensive and integrated context [56]. Proteins with annotated COGs IDs have been included in our study. We have calculated the percentage of each of the COGs categories present in the core proteome of the phylum Deinococcus-Thermus.

RESULTS

(a) Synonymous codon usage bias in radiation-resistant Deinococcus

From Table 1, it is quite evident that the genomes of the phylum *Deinococcus-Thermus*, not only demonstrate a disparity with regard to their radiation-resistance, but they also display a wide

variation in their thermophilic adaptation. For assessing whether the pattern of synonymous codon usage in radiation-resistant *Deinococcus* contains any specific signature, we have performed a correspondence analysis on relative synonymous codon usage (RSCU) of 48, 038 open reading frames (ORFs) from 19 genomes from the microbial phylum *Deinococcus-Thermus* (Listed in Table 1). In Figure 1, Axis1-Axis2 plot of the COA on RSCU has been represented, where the organisms have been segregated on the basis of their thermophilic adaptability along Axis 1.

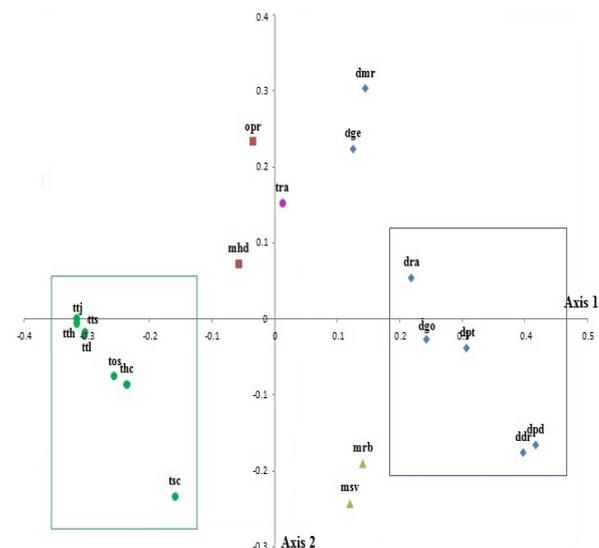


Fig. 1: Correspondence analysis on RSCU. Positions of 19 genomes from the phylum *Deinococcus-Thermus* along first and second principal axes generated by correspondence analysis on relative synonymous codon usage values of 48, 038 ORFs. Mesophilic species (marked by blue box) are present at the extreme positive side of X-axis and hyperthermophilic species (marked by green box) are present at the extreme negative side of X-axis. Thermophilic and slightly thermophilic species are present in between these two extremes.

Mesophilic organisms have their position at the extreme positive side of the X-axis and the hyper-thermophilic ones are present at the extreme negative side of the X-axis. This implies that distinct synonymous codon usage pattern exists for thermophilic and mesophilic organisms under study, as reported by other groups [1, 3, 57, 58].

To obtain further insight, we have analyzed the Codon Adaptation Index (CAI) values for the genomes under study. CAI is an oft used technique for analyzing the codon usage bias [59]. In order to properly understand the CAI, and provide statistical support towards the codon usage bias analysis we have calculated the expected values of Codon Adaptation Index (eCAI) [60] using the server E-CAI (<http://genomes.urv.es/CAIcal/E-CAI>). A significant positive correlation has been observed between the eCAI values and growth temperature in case of the genomes of the phylum *Deinococcus-Thermus* (Spearman's $r_2 = 0.691$, $P < 10^{-2}$) (Supplementary Figure 2), indicating a higher codon usage bias in case of hyperthermophilic order *Thermales* in comparison to mesophilic *Deinococcales*. While differences in thermophilic adaptation explain variation along the first and second major axes (representing 53.16% and 24.8% of total variation, respectively) of the COA of RSCU, the variation along both first and the third major axis separates radiation-sensitive genomes from radiation-resistant genomes. In Figure 2, the Axis1-Axis3 plot of the COA on RSCU has been presented, which shows that the RR and RS genomes, irrespective of their thermophilic adaptiveness have been separated diagonally into two clusters.

have an insight into the amino acid usage behavior of the organisms. Along with the nineteen aforesaid organisms, one mesophilic Actinobacterial species - *Salinispora arenicola* (saq) (GC-content (69.5%) similar to those of the radiation-sensitive hyperthermophilic order *Thermales*), one hyperthermophilic bacteria- *Candidatus Nitrospira defluvii* (nde) (GC-content (59%) less than radiation-sensitive hyperthermophilic order *Thermales*) and one green non-sulfur bacteria - *Thermomicrobium roseum* DSM 5159 (tro) (GC-content (64.3%) similar to the radiation-resistant mesophilic order *Deinococcales*) were also included in our dataset for an extensive comparative analysis.

Actinobacterial and Green non-sulfur bacterial groups have already been reported to be closer to the phylum *Deinococcus-Thermus* [9, 62]. The single-linkage clustering pattern (Figure 5) reveals that the major branching of the organisms is governed by a group specific amino acid usage pattern (Groups have been referred from section 2).

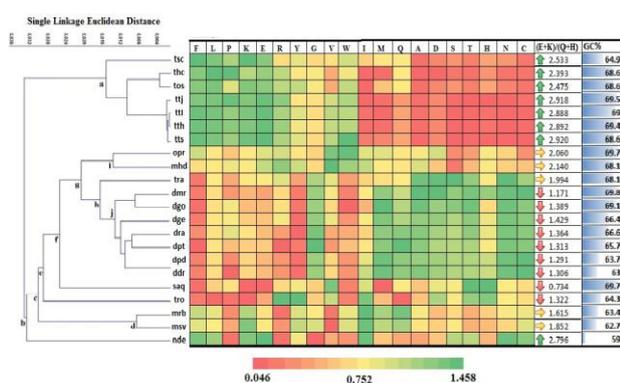


Fig. 5: Single-linkage (Euclidean distances) clustering based on amino acid usages of 19 genomes of the phylum *Deinococcus-Thermus* and three other bacteria (used as reference for comparison). A heat map representing the amino acid usage values has been also provided. The over-representation and under-representations of amino acid residues in the organisms are shown in green and red colored blocks of varying color intensities, respectively.

Table 2: Correlation between the first and third principal axes and nucleotide content in A) Radiation-resistant bacteria, B) Radiation-sensitive bacteria. Correlations with p-value < 10⁻⁴ have been marked in red and correlations with p-value < 10⁻² have been marked in blue. A)

	Axis1	Axis3	GC3	(GC)12	GC
Axis1	1	-0.57	-0.23	-0.09	-0.48
Axis3		1	-0.14	0.08	0.45
GC3			1	-0.06	0.78
(GC)12				1	-0.16
GC					1
B)					
	Axis1	Axis3	GC3	(GC)12	GC
Axis1	1	0.82	-0.62	0.21	-0.61
Axis3		1	-0.72	0.05	-0.68
GC3			1	-0.41	0.99
(GC)12				1	-0.36
GC					1

Moreover, in few cases it has been a cumulative influence of group specific amino acid usage pattern and G+C compositional bias of the organisms. Seven members of the group RS_HT

(Radiation-sensitive, highly thermophilic *Thermus* genomes) cluster within a completely separate node 'a'. Average GC-content of the organisms present within this cluster is 68.35%. It is to be noted that the organisms present in node 'a' show higher usage of the amino acids Lys (K), Glu (E). This result certainly supports the hyper-thermostability of this group of bacteria. According to Farias et al., the E+K/Q+H ratio can be used as an indicator for distinguishing hyperthermophilic, thermophilic and mesophilic bacteria. High percentage of the amino acid couple E+K, along with (The word "couple" has been used twice within a single line) with a low percent of the pair Q+H, influence protein thermostability [63]. In our study, hyperthermophilic bacteria present under node 'a' shows highest values for the E+K/Q+H ratio, supporting its observed hyperthermophilic behavior. The other hyperthermophilic bacteria i.e., *Candidatus Nitrospira defluvii* (nde) (node b), in spite of having much less GC-content than that of the hyperthermophilic *Thermus* group, shows similar value of E+K/Q+H ratio. The preference of group 'a' bacteria towards the usage of amino acids Lys and Glu and avoidance of amino acids Gln (Q) and His (H) is a clear motive for their hyperthermophilic behavior. It is this property that separates them from the rest of the thermophilic and mesophilic bacteria. To be more specific, it can be also noted, that the mesophilic bacteria present under node 'j' shows lowest values for E+K/Q+H ratio and the rest of the bacteria with thermophilic or slightly thermophilic behavior display intermediate values.

It is evident from Figure 5 that node 'b' acts as a point of bifurcation between the hyperthermophilic bacteria 'nde' and the rest of the bacteria under study. Though, 'nde' is a hyperthermophilic bacterium it forms a different clade from node 'a' hyperthermophilic bacteria. In case of few amino acids 'nde' shows diverse usage than the other hyperthermophilic group 'a' bacteria. For example, the usage of GC-rich amino acids Pro, Arg, Gly and Trp, is much less and the usage of GC-poor amino acids Ile and Asn, is much higher in 'nde' than the hyperthermophilic group 'a' bacteria. The comparatively lower GC content of 'nde' accounts for the different clustering from Group 'a' bacteria. This observation can simply be attributed to the GC content factor influencing the amino acid usage pattern. Node 'b' is followed by another bifurcation at node 'c' between the two *Meiothermus* species and the rest of the bacteria. Two *Meiothermus* species (mrb and msv) are clustering together under the node 'd'. These two *Meiothermus* species with comparatively less GC-content (63.05%) represent low-temperature adapted order within *Thermales*, with optimum growing temperature 60°C. Therefore, in spite of being evolutionarily closer to the hyperthermophilic *Thermus* sp., their amino acid usage is closely related to hyperthermophilic 'nde', that has GC-content fairly similar to *Meiothermus* sp. From the foregoing discussion, it can be inferred that the amino acid usage of the two *Meiothermus* sp., clustering under node 'd', has been markedly influenced by their GC-content. Both the above examples reveal the significance of GC content in shaping the amino acid usage patterns of the organisms present in nodes 'b' 'c' and 'd'.

Node 'e' acts as a point of divergence between the green non-sulphur bacteria 'tro' and rest of the eleven bacteria. Node 'e' is followed by another division i.e., node 'f' that separates *Actinobacteria* ('saq') from the phylum *Deinococcus-Thermus*. Mesophilic 'saq' has similar GC-content with the hyperthermophilic group present under node 'a'. But, it has similar amino acid usage pattern like mesophilic bacteria present under node 'j'. It signifies the fact that temperature adaptation influences the amino acid usage pattern of the *Actinobacteria* 'saq'.

Under Node 'g', two thermophilic radiation-sensitive species of the phylum *Deinococcus-Thermus* (opr and mhd) gets completely separated from all the other radiation-resistant *Deinococcus-Thermus* genomes and cluster under node 'i'.

It is quite fascinating to note that node 'h' thermophilic radiation-resistant organism 'tra' clusters with node 'j' organisms representing mesophilic radiation-resistant ones (dge dra dpt dpd and ddr) and

slightly thermophilic radiation-resistant type (dgo and dmr). These groups of organisms reveal somewhat similar amino acid usage patterns. The usage of the aromatic amino acids like Phe, Trp and Tyr is much less, while the usage of small amino acids like Gly, Ala, Ser, Asp, Thr and Asn and sulphur-containing amino acid Cys is much higher in 'tra'. From this perspective, thermophilic, radiation-resistant organism 'tra' not only shows similarity with the thermophilic, radiation-resistant organisms present under node 'j', but also, differ from the radiation-sensitive, thermophilic bacteria 'opr', 'mhd' and the radiation-sensitive, hyperthermophilic bacteria under node 'a'.

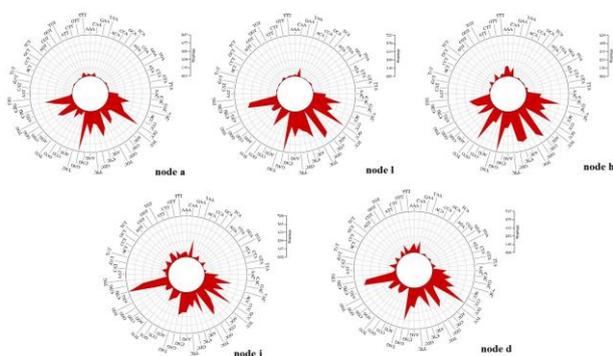


Fig. 6: Rose plot displaying the codon usage pattern for all the 20 amino acids in five separate nodes as symbolized in Fig. 5. The frequencies of all 20 amino acids have been plotted on a radial scale.

This result implies that radiation-adaptation in 'tra' is a crucial determinant in shaping its amino acid usage pattern masking the temperature-adaptation effect. According to Sghaier & Steinar, small and sulphur containing amino acids are preferred and aromatic rings are avoided in proteins of the radiation-resistant bacteria [64]. Therefore, from the above results, it can also be concluded that the organisms, under 'h' show amino acid usage biased on their radiation-adaptation. We have already mentioned that organisms belonging to node 'a' have a tendency to ignore Alanine. It is obvious from the Rose Plots (Figure 6), that the usage of the codon GCC is lowest for the organisms in Group 'a' than all the nodes except node 'h' containing *Truepera radiovictrix*.

It means that avoidance towards the codon GCC has played a crucial role in persuading the organisms belonging to this node for using Alanine in lesser amount than the other nodes. Though Group 'h' also shows avoidance towards the codon GCC, but comparatively higher usage of Alanine by this group can be accounted by the use of alternative Alanine coding codon GCG. Higher usage of amino acid 'D' in node 'h' can be attributed to higher frequency of the codon GAC. Moreover, the codon GAG is used in higher occurrence by the organisms belonging to node 'l' and separate this node from nodes 'j' and 'h' due to variable usage of the amino acid Glutamic acid.

(c) Analyzing the physico-chemical and structural features of RR and RS genomes

We have carried out a comparative analysis of the core proteome i.e., proteins present in all the 19 genomes of the phylum *Deinococcus-Thermus*, to gain a better insight into the physico-chemical properties and structural features helping the genomes to resist radiation as well as thermal stress. Different proteomic properties of 417 orthologous proteins between all the 19 organisms have been summed up in Table 3, which shows a gradual increase in mean hydrophobicity, average pI values, aromaticity and instability indices with the transition from RR group to RS group. In other words, core proteomes of radiation-resistant members (RR) of the phylum *Deinococcus-Thermus* are

less aromatic, less hydrophobic, more acidic and more stable than their radiation-sensitive (RS) orthologs. The comparison of structural properties reveals that irrespective of differential levels of thermophilic adaptation, RR genomes have highest propensities for coil and beta-sheet formation and lowest propensities for helix formation, whereas the RS genomes show a completely opposite trend.

All the RR *Deinococcus* genomes, except the only thermophilic one (tra), reveal similar secondary structural trend. 'tra' (*Truepera radiovictrix*), have higher propensity for helix formation and lesser propensity for coil/beta-sheet formation than the rest of the RR genomes.

RS genomes can as well be grouped according to their thermal adaptations. We have also observed variations in physico-chemical and structural features among these groups. 'msv' and 'mrb' with comparatively low temperature adaptation have highest coil/sheet forming propensity and lowest alpha-helix forming propensity than the rest of the RS members. Their stability is also highest within the RS group. Moreover, RS genomes with high-thermophilic adaptation, show lowest tendency for coil/beta-sheet formation and highest affinity for alpha-helix formation.

These hyper-thermophilic RS genomes experience least stability than the rest of the RS genomes. Thermophilic RS genomes (RS_T) viz., 'opr' and 'mhd', are characterized by intermediate values for all the three structural features. All the above variations in physico-chemical properties as well as structural properties can be a manifestation of their genomic GC-content. To tackle this problem, we have carried out a comparative analysis of all the above properties, within the orthologous sequences from six representatives of the phylum *Deinococcus-Thermus* viz., ddr, dmr, mhd, mrb, tra and ttj (from the groups RR_M, RR_ST, RR_T, RS_LT, RS_T and RS_HT respectively) (for details see Table 1), along with five reference species from other bacterial domains as follows. Three actinobacterial species, viz.,

i) *Salinispora arenicola* CNS-205 (saq): GC-content 69.5% i.e., comparable to that of the radiation-sensitive hyperthermophilic order *Thermales*. 'saq' is also radio-sensitive but mesophilic in nature.

ii) *Rubrobacter xylanophilus* (rxy): 'rxy' is a radiation-resistant and thermophilic bacteria, with GC-content (70.5%) quite comparable with the radiation-resistant, thermophilic bacteria *Truepera*, of the phylum *Deinococcus-Thermus*.

iii) One hyperthermophilic bacteria i.e., *Aquifex aeolicus* (aae) with GC-content 43.32% have been also considered for comparison. In spite of having much lowered GC-content, the hyperthermophilic property of 'aae' is comparable with that of the hyperthermophilic order *Thermales* [10, 65].

Actinobacterial species has been considered for comparison as they are very much close to the phylum *Deinococcus-Thermus* from evolutionary view-point [62]. Values of all the physico-chemical properties and structural propensities of the orthologs from the reference species and the representative *Deinococcus-Thermus* species have been summarized in Table 4. A close look at Table 4 reveals that the hyper-thermophilic bacterial species 'aae', in spite of having much less GC-content, show quite similar aromaticity and structural propensity with that of the radiation-sensitive hyper-thermophilic *Deinococcus-Thermus* species i.e., group RS_HT.

Their stability is also similar with the group RS_HT. On the other hand, 'saq' having comparable GC-content with the group RS_HT, show significantly different values from them, with respect to structural propensities and physico-chemical properties. More interestingly, *Rubrobacter xylanophilus* (rxy), the radiation-resistant *Actinobacterial* species, in spite of having much higher GC-content, show somewhat similar aromaticity and structural trait with radiation-resistant thermophilic group RR_T. The above analyses leads us to conclude that the inter-group

variation in proteomic composition as well as structural traits observed within the phylum *Deinococcus-Thermus* is not a simple signature of their GC-bias, but the influence of other selection

pressures are also considerable, that help them to acclimatize with thermophilic and radiation stress.

Table 3: Various amino acid indices and secondary structural traits of 147 orthologous proteins present in 19 genomes from the phylum *Deinococcus-Thermus*

Organism	Mode of adaptation	Amino acid indices (Mean)			Secondary structural traits (%)			
		Hydrophobicity	Aromaticity	Isoelectric point(pI)	Instability Index (II)	Alpha-helix	Coil	Beta sheet
ddr	RR_M	-0.24	0.060	6.61	37.76	41.22	50.19	8.59
dpt	RR_M	-0.23	0.060	6.41	37.50	41.24	50.43	8.33
dra	RR_M	-0.23	0.061	6.74	37.78	40.3	52.28	7.42
dpd	RR_M	-0.24	0.061	6.72	37.34	40.34	50.77	8.89
dgo	RR_M	-0.22	0.062	6.74	36.54	41.11	50.58	8.31
dmr	RR_ST	-0.21	0.061	6.78	35.54	41.58	50.65	7.77
dge	RR_ST	-0.20	0.060	6.92	36.98	41.72	50.96	7.32
tra	RR_T	-0.20	0.061	6.97	37.85	43.66	49.76	6.58
mrh	RS_LT	-0.18	0.069	7.48	38.75	54.16	39.45	6.39
msv	RS_LT	-0.19	0.070	7.27	38.65	54.61	38.87	6.52
mhd	RS_T	-0.19	0.070	7.29	38.89	56.61	37.87	5.52
opr	RS_T	-0.18	0.069	7.17	38.68	59.3	35.37	5.33
thc	RS_HT	-0.17	0.071	7.33	38.74	62.43	32.25	5.32
tsc	RS_HT	-0.17	0.071	7.32	38.47	61.95	32.52	5.53
tth	RS_HT	-0.18	0.071	7.41	38.25	61.31	33.41	5.28
ttj	RS_HT	-0.18	0.071	7.34	38.16	62.06	32.56	5.38
tth	RS_HT	-0.18	0.071	7.32	39.22	62.66	31.88	5.46
tts	RS_HT	-0.18	0.071	7.36	39.52	62.98	31.44	5.58
tos	RS_HT	-0.18	0.071	7.39	39.58	62.27	32.38	5.35

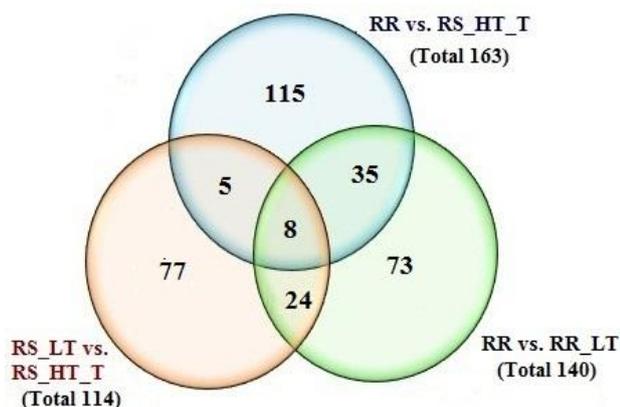


Fig. 7: Venn diagram representing the number of positively selected ($d_N/d_S > 1$) orthologs between pairs of different *Deinococcus-Thermus* genomes. The 147 core proteins of 3 groups (RR, RS_LT, and RS_HT_T) were considered for analysis.

(d) Orthologs from radiation-resistant and radiation-sensitive organisms experience higher positive selection

From the above results it is evident that two groups of genomes from the phylum *Deinococcus-Thermus* i.e., *Deinococcales* and *Thermales* have been found to contain different genomic signatures at the nucleotide as well as amino acid level. Genomic sizes of these two groups are also highly variable. Within the radiation-resistant (RR) *Deinococcales* order, variation in thermophilic adaptation again forms three separate groups; RR_M i.e., mesophilic, RR_ST i.e., slightly thermophilic and RR_T i.e., thermophilic. In spite of having small differences in their

genomic signature, all these three groups have similar genomic sizes. On the other hand, radiation-sensitive (RS) order *Thermales* can also be separated into three groups according to their thermophilic adaptation; RS_LT i.e., thermophilic but adapted in comparatively low temperature range, RS_T i.e., thermophilic and RS_HT i.e., highly-thermophilic. Among these three groups RS_LT has higher genomic size than the rest two groups. Considering the above facts, here we are focusing towards understanding the evolutionary drift within three groups viz., i) RR (includes RR_M, RR_ST and RR_T), ii) RS_LT and iii) RS_HT_T (includes RS_HT and RS_T). The rates of synonymous (d_S) and non-synonymous substitution (d_N) have been calculated between 417 orthologs from RR, RS_LT and RS_HT_T groups. Then, we determined the number of positively selected genes with $d_N > d_S$ values between each possible pair of organisms. In Figure 7, a Venn diagram was drawn for the number of positively selected genes among the genomes under study.

Maximum number of positively selected genes (115) has been observed between the orthologs from the groups RR and group RS_HT_T; organisms from these two groups differ in genomic size and radiation adaptation. The groups RS_LT and RS_HT_T, with similar trend of radiation adaptation but different genomic size, stand subsequently with 77 positively selected genes among their orthologs. The groups RR and RS_LT with similar genomic size but dissimilar radiation-adaptation contain minimum number of positively selected genes (73) among their orthologs. From the Venn diagram, it is also observed that there are 35 positively selected genes between the group RR and either of the RS groups i.e., between the groups of two opposite radiation-optima, irrespective of genomic size. All the above observations concludes that random genetic drift plays appreciable role in vertical niche partitioning of the genomes according to their adaptation in different radiation optima as well as thermophilic condition [66]

Table 4: Comparison between various amino acid indices and secondary structural traits of five sets of orthologous proteins from the phylum *Deinococcus-Thermus* and five other reference bacteria

Organisms		Amino acid indices (Mean)			Secondary structural traits (%)			
		Hydrophobicity	Aromaticity	Isoelectric point(pI)	Instability Index (II)	Alpha-helix	Coil	Beta sheet
set I (370 pairs)	aae	-0.28	0.08	7.75	37.76	52.32	34.95	10.73
	RR_M	-0.25	0.06	7.49	34.81	40.83	46.63	12.55
	RR_ST	-0.18	0.06	7.44	32.88	42.75	44.98	12.26
	RR_T	-0.18	0.06	7.73	35.73	45.11	42.97	11.92
	RS_LT	-0.18	0.06	7.88	34.94	46.65	40.41	12.94
	RS_T	-0.21	0.06	7.78	36.24	49.71	37.41	11.88
set II (388 pair)	RS_HT	-0.18	0.08	7.82	37.53	53.11	34.30	10.78
	saq	-0.16	0.04	6.36	32.68	38.07	47.70	13.92
	RR_M	-0.25	0.06	7.50	34.35	41.20	46.58	12.21
	RR_ST	-0.23	0.06	7.64	32.36	40.86	47.28	11.86
	RR_T	-0.20	0.06	7.92	36.31	47.13	43.49	9.39
	RS_LT	-0.21	0.06	7.94	35.00	48.54	41.34	10.11
set III (380 pairs)	RS_T	-0.24	0.06	7.75	35.43	53.07	36.92	10.01
	RS_HT	-0.22	0.06	7.85	35.69	54.38	35.57	10.05
	rxxy	-0.18	0.06	7.21	36.49	42.82	42.05	9.98
	RR_M	-0.21	0.06	7.27	33.24	46.53	47.46	10.01
	RR_ST	-0.20	0.06	7.23	32.95	44.93	45.16	9.91
	RR_T	-0.18	0.06	7.13	36.80	42.50	42.37	9.07
set III (380 pairs)	RS_LT	-0.19	0.06	7.88	38.25	48.46	41.49	8.89
	RS_T	-0.21	0.07	7.56	34.01	48.90	41.83	8.26
	RS_HT	-0.21	0.07	7.73	35.96	54.85	39.43	8.63

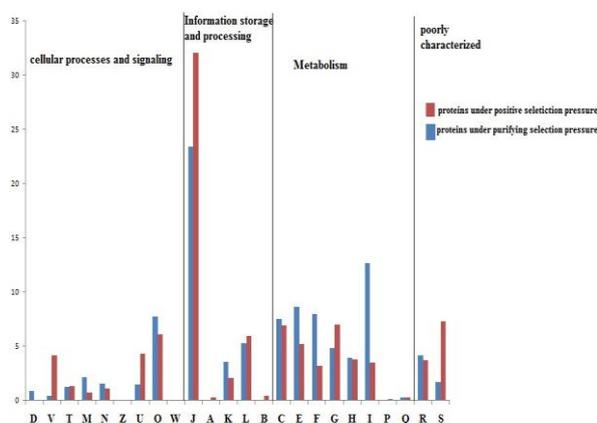


Fig. 8: Bar-diagram representing distribution of different COG categories for the core-proteome of the phylum *Deinococcus-Thermus*. Description of COG categories are as follows: [J] Translation, ribosomal structure and biogenesis, [A] RNA processing and modification, [K] Transcription, [L] Replication, recombination and repair, [B] Chromatin structure and dynamics, [D] Cell cycle control, cell division, chromosome partitioning, [Y] Nuclear structure, [V] Defense mechanisms, [T] Signal transduction mechanisms, [M] Cell wall/membrane/envelope biogenesis, [N] Cell motility, [Z] Cytoskeleton, [W] Extracellular structures, [U] Intracellular trafficking, secretion, and vesicular transport, [O] Posttranslational modification, protein turnover, chaperones, [C] Energy production and conversion, [G] Carbohydrate transport and metabolism, [E] Amino acid transport and metabolism, [F] Nucleotide transport and metabolism, [H] Coenzyme transport and metabolism, [I] Lipid transport and metabolism, [P] Inorganic ion transport and metabolism, [Q] Secondary metabolites biosynthesis, transport and catabolism, [R] General function prediction only, [S] Function unknown.

(e) COG classification of the core proteome of the phylum *Deinococcus-Thermus*

We have divided the 417 core proteins from each of the 19 genomes into two categories according to the selection pressure acting on them i.e., proteins under purifying selection pressure ($d_N < d_S$) and proteins under positive selection pressure ($d_N > d_S$). Then, we have used COGs functional classification to compare these two categories of proteins. Figure 8 represents the comparison of COG classification of proteins under purifying selection and positive selection.

From the comparison, one can observe that proteins having function "Information Storage and Processing" mostly experience positive selection pressure. Besides this, proteins under the COG category V (Defense Mechanisms) and U (Intracellular Trafficking, Secretion, and Vesicular Transport) are also under positive selection pressure. This result indicates that proteins belonging to the above COG categories are accumulating non-synonymous mutations during the vertical evolutionary process within the phylum *Deinococcus-Thermus*, facilitating its members to adapt in their specific niche area.

On the other hand, purifying selection pressure is acting upon the proteins performing the function "Metabolism", along with the proteins under the COG category M (Cell wall/Membrane/Envelope Biogenesis), N (Cell Motility) and O (Post-translational Modification, Protein Turnover, Chaperones), confirming the conservation of these COG categories throughout the evolutionary drift within the phylum *Deinococcus-Thermus*.

DISCUSSION

Differences in genomic G+C content (62.7% to 69.8%) and genome sizes (2.1 Mb to 4.5 Mb) among the various genomes of the *Deinococcus-Thermus* phylum, along with their adaptation in radiation and thermophilic stressed conditions require a detailed investigation. In the present study, we have performed a large-scale comparison at the genomic as well as proteomic level to figure out the direction and strength of mutational pressure associated with the evolution of the organisms present within the phylum *Deinococcus-Thermus*, which may shed light into the intricacies of molecular evolution. From our study, few interpretations can be

inferred that support the molecular evolution of the genomes of the phylum *Deinococcus-Thermus* for adapting in their particular habitat.

(a) Present study indicates radical changes in the patterns of synonymous codon usage bias between the hyperthermophilic *Thermales* genomes and mesophilic *Deinococcales* genomes. Our observation can further be supported by the work of Lynn et al. which has already established that particular environmental factors, such as high temperature in eubacterial lineages, can inspire natural selection forces that govern the synonymous codon usage patterns in genomes. [57]. Besides this, there is a variation in the synonymous codon choices between the radiation-resistant and radiation-sensitive *Deinococcus-Thermus* genomes. Preference towards 'G'- or 'C'-ending codons is quite higher in case of the order *Thermales* than that of the order *Deinococcales*. From the present work, it can be concluded that during the evolutionary process, horizontal acquisition of few explicit genes with high G+C content by the radiation-sensitive hyperthermophilic *Thermales* genomes, may account for the above mentioned variation in favor of 'G'- or 'C'-ending codons between the order *Thermales* and *Deinococcales*. Omelchenko et al., observed that the common ancestor of *Deinococcales* and *Thermales* group of organisms was a mesophilic bacterium. Comparative genomics and phylogenetic analysis had revealed the presence of a few additional "thermophilic determinant" genes in the genome of 'dge' (a slightly thermophilic radiation-resistant *Deinococcus* genome). Notably, those genes are missing in mesophilic radiation-resistant *Deinococcus* genomes (Table 1). It was also suggested earlier that few proteins have been acquired horizontally by the hyperthermophilic radiation-sensitive *Thermus* genomes from highly thermophilic bacteria like *Aquifex* [10]. These proteins have been revealed to be lactose utilization proteins (LacZ, GalK, GalT, GalA) and proteins involved in Sox-like sulfur oxidation system viz., SoxA, SoxB, SoxC, SoxD, SoxX, SoxY and SoxZ. Earlier studies have also recommended about proteins ('degradation and amino acid catabolism proteins' 'secreted subtilisin-like proteases' and 'aerobic-type carbon monoxide dehydrogenase') to have been horizontally acquired by the radiation-resistant *Deinococcus* genomes from different mesophilic bacteria [10]. Present analysis depicts a significant influence of mutational pressure that accounts for considerable amount of differences in synonymous codon usage bias between the radiation-resistant and radiation-sensitive genomes of the phylum *Deinococcus-Thermus*. The radiation-resistant order *Deinococcales* and the radiation-sensitive order *Thermales* have followed their specific evolutionary tracks after speciation event from a common mesophilic ancestor. Evolution of the hyperthermophilic phenotype of radiation sensitive (RS) order *Thermales* has taken place by horizontal gene transfer from thermophilic organisms other than the order *Deinococcus-Thermus*. On the other hand, the radiation resistant (RR) phenotype of the order *Deinococcales* might have gradually evolved via horizontal transfer of genes from the ancestral mesophilic organisms, other than the order *Deinococcus-Thermus*, with a highly sophisticated oxidative stress response system [10]. The Horizontal Gene Transfer DataBase (HGT-DB) (<http://genomes.urv.es/HGT-DB/>) [67] provides information about the horizontally acquired genes of 'dra' (radiation-resistant mesophilic genome of the order *Deinococcales*), 'dge' (radiation-resistant slightly thermophilic genome of the order *Deinococcales*), 'ttj' and 'tth' (both are radiation-sensitive highly thermophilic genome of the order *Thermales*). Using this information we have retrieved 83, 101, 91 and 103 coding regions of horizontally transferred genes from the 'dra', 'dge', 'ttj' and 'tth' genomes respectively and calculated their GC-content. The average GC-content of the genes horizontally acquired by highly thermophilic radiation-sensitive genome of the order *Thermales* ('ttj' and 'tth') have been found to be comparatively higher (about 66.3%) than the average GC-content of the genes acquired by the radiation-resistant mesophilic genome of the order *Deinococcales* ('dra') (about 58.2%).

On the other hand, the average GC-content of the genes horizontally acquired by slightly thermophilic radiation-resistant genome of the order *Deinococcales* ('dge') is nearly 62.2%. According to Nishida et al., genetic information is transferred from AT-rich chromosomes to GC-rich chromosomes during evolution [68]. Therefore, from the present work it might be said that the radiation-sensitive

hyperthermophilic order *Thermales*, had to increase their genomic GC-content to horizontally acquire the thermophilic genes having a higher G+C content. As the selection pressure is comparatively much less in the third codon position [69-73], radiation-sensitive hyperthermophilic order *Thermales* have increased their tendency of preferring the 'G'- or 'C'-ending codons to enhance their overall genomic G+C content. Nearly seven-fold higher significant negative correlation of (G+C) 12% and (G+C) 3% in case of radiation-sensitive hyperthermophilic order *Thermales*, as revealed by the present study, supports the above fact.

(b) Wu et al., have postulated that increased bacterial genomic size appears to rely on increased genomic GC content [74]. However, in the current study we have noticed that the genomic sizes and growth temperature of the 19 members belonging to the phylum *Deinococcus-Thermus*, share significantly negative correlation (Spearman's $r_2 = -0.852$, $P < 10^{-2}$), while genomic size and GC-content show significant positive correlation with each other (Spearman's, $r_2 = 0.42$, $P < 10^{-2}$). Intergenic regions are parts of genomes that are under least constraint. Therefore, if selection plays role towards decreasing the genomic size, then, the reduction in size should preferentially affect the intergenic regions of a genome. However, this is not a case as is evident for bacterial strains [75-77]. To elucidate this criterion, in our case, we calculated the percentage of DNA content in intergenic regions and found that with the increase of growth temperature, the percentage of intergenic regions have significantly decreased (Spearman's, $r_2 = -0.843$, $P < 10^{-2}$) (Supplementary Figure 4) for the genomes under study. According to Sakharkar et al., the reduction of intergenic regions gives rise to gene overlaps that finally help in genomic size reduction while preserving the information content [78].

Therefore, the percentage of overlapping genes in all the 19 genomes has also been calculated. Overlapping genes can be defined as pairs of adjacent genes whose coding regions partly or completely overlap. Coding sequence (CDS) annotation features were used for extracting the genes showing overlap. It has been found that the percentage of overlapping genes have significantly increased (Spearman's, $r_2 = 0.876$, $P < 10^{-2}$) with the increase of growth temperature (Supplementary Figure 5) signifying a clear trend of genome reduction strategy. As the members of the order *Deinococcales* are mostly mesophilic and the members of the order *Thermales* are mostly hyperthermophilic, the above findings indicate that in spite of having a comparatively lesser (G+C) content, mesophilic *Deinococcales* genomes have higher genomic size than hyperthermophilic *Thermales* genomes, due to the significant reduction in the percentage of overlapping genes and amplification in the percentage of DNA content in the intergenic regions.

It has been also found earlier that during their different evolutionary patterning, considerable number of gene-flux had taken place in both the order *Deinococcales* and *Thermales*. But, the total gene-gain through horizontal gene transfer was quite higher in radiation-resistant, mesophilic *Deinococcales* genomes than that of radiation-sensitive, hyperthermophilic *Thermales* genomes [10], facilitating the increase of genomic size in *Deinococcales* genomes than *Thermales* genomes. All the above facts support the idea that after getting separated from their mesophilic common ancestor, the evolutions of radiation-resistant, mesophilic *Deinococcales* genomes and radiation-sensitive, hyperthermophilic *Thermales* genomes have occurred discretely.

(c) Amino acid usage pattern of eight radiation-resistant organisms under node 'h' has been influenced by their mode of adaptation in radiation rather than their GC-content. But, the amino acid usage pattern of the nine radiation-sensitive, hyperthermophilic organisms under node 'a' and 'd' is governed by their GC-content instead of their mode of adaptation. Moreover, the amino acid usage pattern of two radiation-sensitive thermophilic bacteria under node 'i' is a reflection of their thermophilic adaptation. It is noteworthy that the differences in the pattern of amino acid usage as observed, in our analysis, between hyperthermophilic radiation-sensitive *Thermales* genomes and mesophilic radiation-resistant *Deinococcales* genomes are quite compatible with an earlier report by John H. McDonald [39], where he has mentioned similar pattern

of differences in amino acid usage between radiation-resistant, mesophilic bacteria *Deinococcus radiodurans* and radiation-sensitive, hyperthermophilic bacteria *Thermus thermophilus*, facilitating their temperature adaptation. Higher Lys (K), Glu (E) content of the hyperthermophilic order *Thermales* along with elevated values of [Lys (K) + Glu (E)] / [Gln (Q) + His (H)] ratio than the mesophilic and slightly thermophilic *Deinococcales* genomes, as observed in the present study, clearly supports the separate course of evolution of the order *Thermales* to adapt in high thermophilic conditions [63]. On the other hand, preference towards the usage of sulphur containing small amino acids and avoidance of aromatic amino acids in the proteins of the radiation-resistant *Deinococcus* genomes, sustains their separate evolutionary trace to adapt in radiation [64].

(d) Core proteome of different groups of the phylum *Deinococcus-Thermus* species show variation in different physico-chemical and structural features. Higher hydrophilicity within the core proteome of radiation-resistant *Deinococcales* genomes is also supported by the work by Misra et al., (2013) [79]. Some proteins of *D. radiodurans*, involved in cellular recovery processes (for radiation-resistance) and desiccation-resistance have been reported to possess large hydrophilic regions [79, 80]. Presence of higher proportion of coiled structure and lesser amount of helix structures in the core proteome of radiation-resistant *Deinococcales* genomes as detected in the present study tends to be more flexible. Consequently, they are more stable than their orthologs from radiation-sensitive, hyperthermophilic, *Thermales* genomes. Organisms belonging to the radiation-resistant *Deinococcales* genomes, would adopt to withstand ionizing radiation and UV radiation (100 to 295 nm), as well as desiccation, which induce oxidative damage not only to DNA but also to all cellular macromolecules via the production of reactive oxygen species. Therefore, extreme flexibility of the radiation-resistant *Deinococcales* genomes is beneficial for their powerful DNA repair machinery and an efficient protection of proteins against oxidative stress [19, 81]. On the other hand, in addition to the intrinsic helix forming propensities, presence of more hydrophobic and aromatic residues within the core proteome of the radiation-sensitive thermophilic *Thermales* genomes, facilitate their thermophilic adaptation [42]. Besides this, higher pI values for these hyperthermophilic radiation-sensitive *Thermales* genomes might facilitate cation-pi interaction in them leading to their thermal stability [82]. Thus, we may conclude that modification of amino acid composition that could potentially affect protein stability is another crucial determinant that could contribute to the genome differentiation of RR and RS organisms from their common ancestor.

Besides this, we have also observed that:

(e) Maximum number of positively selected genes is present between the orthologs from the organisms with distinct genomic size and radiation optima i.e., radiation-resistant *Deinococcales* genomes and radiation-sensitive *Thermales* genomes, which points towards a random genetic drift for vertical niche partitioning of the phylum *Deinococcus-Thermus* into two orders i.e., *Deinococcales* and *Thermales* according to their adaptation in different radiation optima as well as thermophilic condition.

(f) Core proteins of the phylum *Deinococcus-Thermus* associated with functions such as defense mechanisms, intracellular trafficking, secretion, and vesicular transport mostly experience positive selection pressure. The mentioned genes fall under the COG categories: a) CELLULAR PROCESSES AND SIGNALING and b) INFORMATION STORAGE AND PROCESSING. This observation signifies the fact that during the vertical evolutionary process within the phylum *Deinococcus-Thermus*, there has been a pronounced impact of non-synonymous mutational pressure on these groups of proteins. This is clear evidence revealing the adaptive strategy employed by the phylum *Deinococcus-Thermus* to withstand the extremities of temperature and radiation and thus, successfully adapt to their specific niche area (Figure 9).

Core proteins of the phylum *Deinococcus-Thermus* performing functions associated with Defense Mechanisms, Intracellular Trafficking, Secretion, and Vesicular Transport "Information Storage and Processing" mostly experience positive selection pressure. This result reveals the fact that during the vertical evolutionary process within the phylum *Deinococcus-Thermus*, maximum quantity of non-

synonymous mutations have been accumulated within the core proteins responsible for the, facilitating the members of the phylum *Deinococcus-Thermus* to adapt in their specific niche area. All the above results explicitly support a systematic evolutionary pattern of the phylum *Deinococcus-Thermus* genomes during their adaptive evolution (Figure 9).

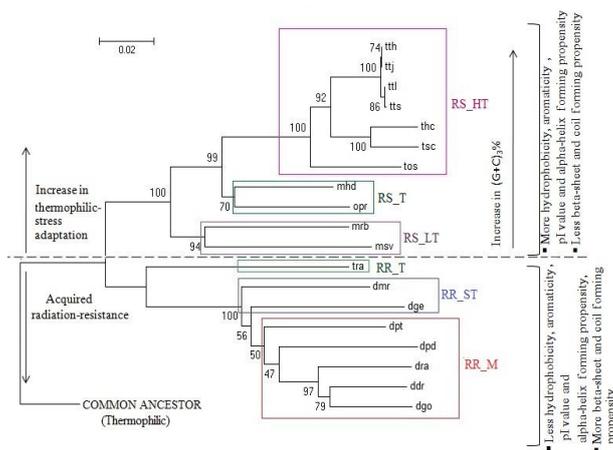


Fig. 9: Evolution of the phylum *Deinococcus-Thermus* at the genomic as well as proteomic level, describes a step-wise diversification of genomes according to their mode of adaptation. The phylogenetic model is based on the 16s rRNA phylogeny of the 19 *Deinococcus-Thermus* genomes, considering the bootstrap consensus tree (500 replicates) generated using Minimum Evolution method of the software MEGA (version 5).

Supplementary

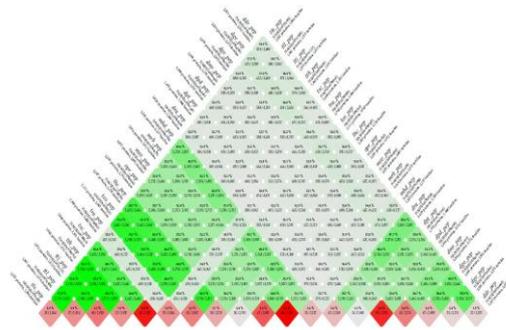


Fig. 1: BLAST Matrix for the 19 genomes. The strongest green represents the highest fraction of genes shared between two genomes under scrutiny. The intensity of colors from green to red suggests the percentage of genes being shared between genomes in a descending order.

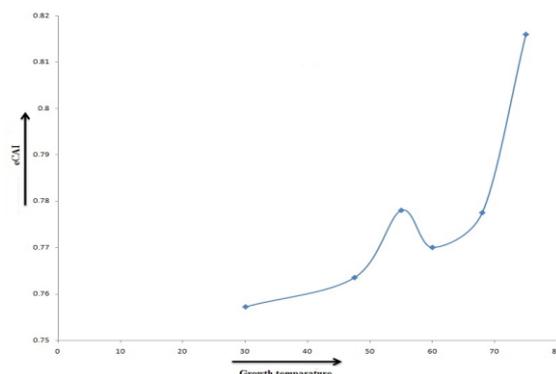
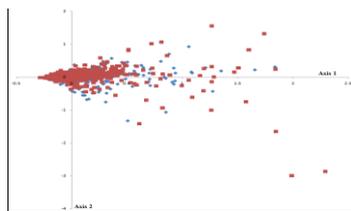
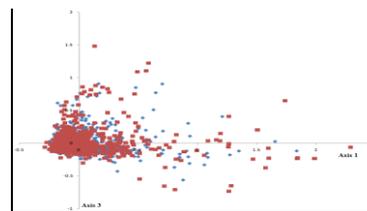


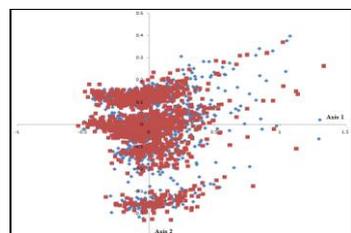
Fig. 2: Graphical representation of eCAI value being plotted against growth temperature for the phylum *Deinococcus-Thermus*.



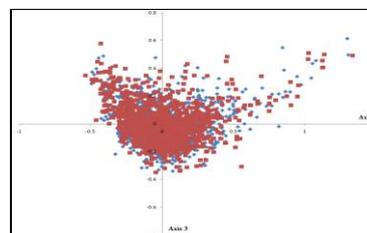
A. Position of genes on the planes defined by the first (horizontal Axis 1) and second (Vertical Axis 2) major axes generated by COA on RSCU values of coding region of ttj (representative hyperthermophilic, radiation-sensitive strain).



B) Position of genes on the planes defined by the first (horizontal Axis 1) and third (Vertical Axis 3) major axes generated by COA on RSCU values of coding region of ttj (representative hyperthermophilic, radiation-sensitive strain).



C) Position of genes on the planes defined by the first (horizontal Axis 1) and second (Vertical Axis 2) major axes generated by COA on RSCU values of coding region of ddr (representative mesophilic, radiation-resistant strain).



D) Position of genes on the planes defined by the first (horizontal Axis 1) and third (Vertical Axis 3) major axes generated by COA on RSCU values of coding region of ddr (representative mesophilic, radiation-resistant strain).

Fig. 3

Genes transcribed from the leading and lagging strands are represented by red and blue colored dots respectively.

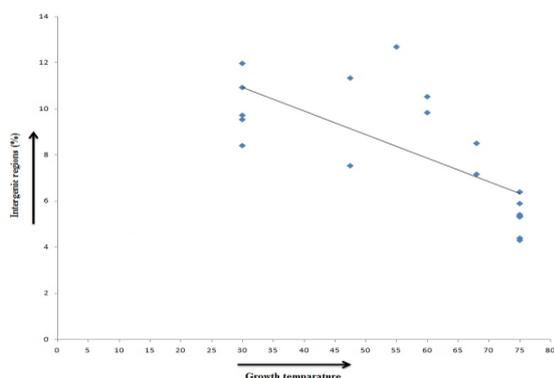


Fig. 4: Graphical representation of percentage of DNA-content of intergenic regions being plotted against growth temperature for the phylum *Deinococcus-Thermus*.

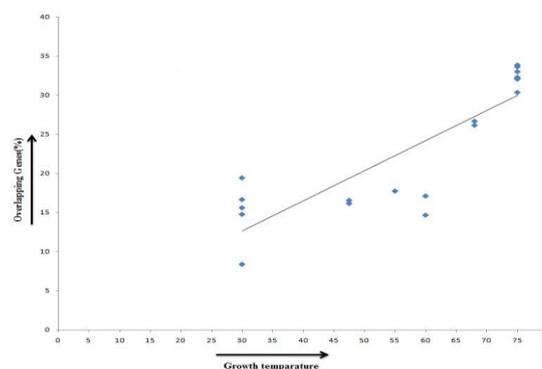


Fig. 5: Graphical representation of percentage of overlapping genes being plotted against growth temperature for the phylum *Deinococcus-Thermus*.

Table 1: Comparison of codon usage values of genes from both radiation-sensitive and radiation-resistant bacterial groups. Codons, usage of which are significantly higher in radiation-resistant group have been marked by blue, while that is higher in radiation-resistant group have been marked by red.

		RR	RS			RR	RS			RR	RS			RR	RS
Phe	UUU	0.60	0.48	Ser	UCU	0.23	0.17	Tyr	UAU	0.33	0.21	Cys	UGU	0.23	0.11
	UUC	1.35	1.48		UCC	0.86	2.11		UAC	1.60	1.75		UGC	1.12	0.96
Leu	UUA	0.68	0.67	Pro	UCA	0.20	0.10	His	CAU	0.35	0.21	Trp	UGG	0.88	0.90
	UUG	0.36	0.55		UCG	1.22	0.96		CAC	1.54	1.65				
	CUU	0.28	0.63		CCU	0.36	0.34		Gln	CAA	0.26		0.33	Arg	CGU
	CUC	1.63	2.29	CCC	1.79	2.55		CAG	1.72	1.62		CGC	3.46	2.42	
	CUA	0.10	0.20	CCA	0.23	0.15						CGA	0.16	0.17	
	CUG	3.57	2.24	CCG	1.61	0.95	Asn	AAU	0.31	0.13		CGG	1.48	2.29	
Ile	AUU	0.67	0.35	Thr	ACU	0.22	0.11	Lys	AAC	1.60	1.71	Ser	AGU	0.51	0.14
	AUC	2.17	2.39		ACC	2.27	2.85		AAA	0.54	0.34		AGC	2.96	2.44
	AUA	0.08	0.15		ACA	0.18	0.08		AAG	1.32	1.62		Arg	AGA	0.09
Met	AUG	0.97	0.96	ACG	1.32	0.91	Asp	GAU	0.29	0.32		AGG	0.19	0.77	
Val	GUU	0.23	0.23	Ala	GCU	0.32	0.24	Glu	GAC	1.70	1.65	Gly	GGU	0.42	0.25
	GUC	1.36	1.15		GCC	1.92	2.59		GAA	0.72	0.40		GGC	2.36	1.70
	GUA	0.17	0.21		GCA	0.26	0.14		GAG	1.26	1.59		GGA	0.29	0.30
	GUG	2.25	2.40		GCG	1.50	1.03						GGG	0.93	1.75

CONCLUSIONS

Present study reveals the fact that G+C compositional constraint has been a major factor that governs codon usage patterns of the phylum *Deinococcus-Thermus*. At the genomic level, synonymous codon usage bias has been found to play a crucial role in segregating the genomes according to their mode of adaptation. On the other hand, at the proteomic level, genomes of the phylum *Deinococcus-Thermus* have revealed distinct amino-acid compositional feature, physicochemical and structural trait and variations in the degree of stability of their core proteome, according to their adaptation towards radiation and variable temperature stresses. Amino acid usage patterns are inferred to be shaped by the combinatorial effect of G+C compositional bias, thermophilic and radiation resistant tendencies. Evolutionary analysis sheds light into the fact that the house keeping genes and the genes falling under the Metabolism COG category are under strong purifying selection pressure which is obvious and expected. However, the most astonishing detail that is revealed from the analysis is that the genes falling under the COG categories- Defense Mechanisms and Intracellular Trafficking, Secretion, and Vesicular Transport are under the influence of positive selection. This reflects the adaptive strategy employed by these pools of microbes to acclimatize to the adversaries in their niche and thus safeguard their existence.

ABBREVIATIONS

RR=Radiation-resistant

RS=Radiation-sensitive

HT=Hyperthermophilic

T=Thermophilic

M=Mesophilic

ST=Slightly Thermophilic

LT= Thermophilic but comparatively low temperature adapted

COA= Correspondence analysis (COA)

RSCU= Relative synonymous codon usage

ACKNOWLEDGEMENT

We wish to thank Department of Biotechnology, Government of India (Grant no.BT/B1/04/026/93 and BT/B1/010/019/99), and University Grants Commission, Government of India (Grant no.UGC/1197B/RFSMS/Biophy) for providing us financial support.

CONFLICT OF INTEREST STATEMENT

There is no conflict of interest among the authors.

REFERENCES

- Das S, Paul S, Bag SK, Dutta C. Analysis of Nanoarchaeum equitans genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* 2006; 7: 186.
- Das S, Paul S, Chatterjee S, Dutta C. Codon and amino acid usage in two major human pathogens of genus Bartonella--optimization between replicational-transcriptional selection, translational control and cost minimization. *DNA Res* 2005; 12: 91-102.
- Das S, Paul S, Dutta C. Evolutionary constraints on codon and amino acid usage in two strains of human pathogenic actinobacteria *Tropheryma whippelii*. *J Mol Evol* 2006; 62: 645-658.
- Dutta C, Paul S. Microbial lifestyle and genome signatures. *Curr Genomics* 2012; 13: 153-162.
- Eisenberg H. Life in unusual environments: progress in understanding the structure and function of enzymes from extreme halophilic bacteria. *Arch Biochem Biophys* 1995; 318: 1-5.
- Paul S, Dutta A, Bag SK, Das S, Dutta C. Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. *BMC Genomics* 2010; 11: 103.
- Griffiths E, Gupta RS. Identification of signature proteins that are distinctive of the *Deinococcus-Thermus* phylum. *Int Microbiol* 2007; 10: 201-208.
- Gupta RS. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 1998; 62: 1435-1491.
- Gupta RS, Bustard K, Falah M, Singh D. Sequencing of heat shock protein 70 (DnaK) homologs from *Deinococcus proteolyticus* and *Thermomicrobium roseum* and their integration in a protein-based phylogeny of prokaryotes. *J Bacteriol* 1997; 179: 345-357.
- Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Daly MJ, Koonin EV, Makarova KS. Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. *BMC Evol Biol* 2005; 5: 57.
- Rainey FA, Nobre MF, Schumann P, Stackebrandt E, da Costa MS. Phylogenetic diversity of the deinococci as determined by 16S ribosomal DNA sequence comparison. *Int J Syst Bacteriol* 1997; 47: 510-514.
- Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Venkateswaran A, Hess M, Omelchenko MV, Kostandarites HM, Makarova KS, Wackett LP, Fredrickson JK, Ghosal D. Accumulation of Mn(II) in *Deinococcus radiodurans* facilitates gamma-radiation resistance. *Science* 2004; 306: 1025-1028.
- Gutman PD, Fuchs P, Minton KW. Restoration of the DNA damage resistance of *Deinococcus radiodurans* DNA polymerase mutants by *Escherichia coli* DNA polymerase I and Klenow fragment. *Mutat Res* 1994; 314: 87-97.
- Lai WA, Kampfer P, Arun AB, Shen FT, Huber B, Rekha PD, Young CC. *Deinococcus ficus* sp. nov., isolated from the rhizosphere of *Ficus religiosa* L. *Int J Syst Evol Microbiol* 2006; 56: 787-791.
- Burrell AD, Feldschreiber P, Dean CJ. DNA-membrane association and the repair of double breaks in x-irradiated *Micrococcus radiodurans*. *Biochim Biophys Acta* 1971; 247: 38-53.
- Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Leapman RD, Lai B, Ravel B, Li SM, Kemner KM, Fredrickson JK. Protein oxidation implicated as the primary determinant of bacterial radioresistance. *PLoS Biol* 2007; 5: e92.
- Dean CJ, Feldschreiber P, Lett JT. Repair of x-ray damage to the deoxyribonucleic acid in *Micrococcus radiodurans*. *Nature* 1966; 209: 49-52.
- Kitayama S, Matsuyama A. Possibility of the repair of double-strand scissions in *Micrococcus radiodurans* DNA caused by gamma-rays. *Biochem Biophys Res Commun* 1968; 33: 418-422.
- Slade D, Radman M. Oxidative stress resistance in *Deinococcus radiodurans*. *Microbiol Mol Biol Rev* 2011; 75: 133-191.
- Blasius M, Sommer S, Hubscher U. *Deinococcus radiodurans*: what belongs to the survival kit? *Crit Rev Biochem Mol Biol* 2008; 43: 221-238.
- Cox MM, Battista JR. *Deinococcus radiodurans* - the consummate survivor. *Nat Rev Microbiol* 2005; 3: 882-892.
- Daly MJ. A new perspective on radiation resistance based on *Deinococcus radiodurans*. *Nat Rev Microbiol* 2009; 7: 237-245.
- Ferreira AC NM, Rainey FA, Silva MT, Wait R, Burghardt J, Chung AP, da Costa MS. *Deinococcus geothermalis* sp. nov. and *Deinococcus murrayi* sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs. *Int J Syst Bacteriol* 1997; 47(4): 939-947.
- Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, Koonin EV, Daly MJ. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol Mol Biol Rev* 2001; 65: 44-79.
- Yuan M, Zhang W, Dai S, Wu J, Wang Y, Tao T, Chen M, Lin M. *Deinococcus gobiensis* sp. nov., an extremely radiation-resistant bacterium. *Int J Syst Evol Microbiol* 2009; 59 (Pt 6): 1513-1517.
- Albuquerque L, Simoes C, Nobre MF, Pino NM, Battista JR, Silva MT, Rainey FA, da Costa MS. *Trueperia radiovictrix* gen. nov., sp. nov., a new radiation resistant species and the proposal of *Trueperaceae* fam. nov. *FEMS Microbiol Lett* 2005; 247: 161-169.

27. Tindall BJ, Sikorski J, Lucas S, Goltsman E, Copeland A, Glavina Del Rio T, Nolan M, Tice H, Cheng JF, Han C, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Ovchinnikova G, Pati A, Fahrnich R, Goodwin L, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Rohde M, Goker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Lapidus A. Complete genome sequence of *Meiothermus ruber* type strain (21). *Stand Genomic Sci* 2010; 3: 26-36.
28. Miroshnichenko ML, L'Haridon S, Jeanthon C, Antipov AN, Kostrikin NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E, Bonch-Osmolovskaya EA. *Oceanithermus profundus* gen. nov., sp. nov., a thermophilic, microaerophilic, facultatively chemolithoheterotrophic bacterium from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 2003; 53: 747-752.
29. Sako Y, Nakagawa S, Takai K, Horikoshi K. *Marinithermus hydrothermalis* gen. nov., sp. nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* 2003; 53: 59-65.
30. Hedlund BP, McDonald AI, Lam J, Dodsworth JA, Brown JR, Hungate BA. Potential role of *Thermus thermophilus* and *T. oshimai* in high rates of nitrous oxide (N₂O) production in approximately 80 degrees C hot springs in the US Great Basin. *Geobiol* 2011; 9: 471-480.
31. Kristjánsson JK, Hjörleifsdóttir S, Marteinson V, Alfredsson GA. *Thermus scotoductus*, sp. nov., a pigment-producing thermophilic bacterium from hot tap water in Iceland and including *Thermus* sp. X-1. *Syst Appl Microbiol* 1994; 17: 44-50.
32. Murugapiran SK, Huntemann M, Wei CL, Han J, Detter JC, Han CS, Erkkila TH, Teshima H, Chen A, Kyrpides N, Mavrommatis K, Markowitz V, Szeto E, Ivanova N, Pagani I, Lam J, McDonald AI, Dodsworth JA, Pati A, Goodwin L, Peters L, Pitluck S, Woyke T, Hedlund BP. Whole Genome Sequencing of *Thermus oshimai* JL-2 and *Thermus thermophilus* JL-18, Incomplete Denitrifiers from the United States Great Basin. *Genome Announc* 2013; 1: e00106-12.
33. Yoshida T, Lorence RM, Choc MG, Tarr GE, Findling KL, Fee JA. Respiratory proteins from the extremely thermophilic aerobic bacterium, *Thermus thermophilus*. Purification procedures for cytochromes c552, c555,549, and c1aa3 and chemical evidence for a single subunit cytochrome aa3. *J Biol Chem* 1984; 259: 112-123.
34. Forterre P. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* 2002; 18: 236-237.
35. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* 2002; 30: 482-496.
36. Makarova KS, Wolf YI, Koonin EV. Potential genomic determinants of hyperthermophily. *Trends Genet* 2003; 19: 172-176.
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. Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc Natl Acad Sci USA* 1999; 96: 3578-3583.
39. McDonald JH. Patterns of temperature adaptation in proteins from the bacteria *Deinococcus radiodurans* and *Thermus thermophilus*. *Mol Biol Evol* 2001; 18: 741-749.
40. Sterner R, Liebl W. Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* 2001; 36: 39-106.
41. Thompson MJ, Eisenberg D. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol* 1999; 290: 595-604.
42. Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 2001; 65: 1-43.
43. Banerjee R, Roy A, Ahmad F, Das S, Basak S. Evolutionary patterning of hemagglutinin gene sequence of 2009 H1N1 pandemic. *J Biomol Struct Dyn* 2012; 29: 733-742.
44. Basak S, Banerjee T, Gupta SK, Ghosh TC. Investigation on the causes of codon and amino acid usages variation between thermophilic *Aquifex aeolicus* and mesophilic *Bacillus subtilis*. *J Biomol Struct Dyn* 2004; 22: 205-214.
45. Basak S, Ghosh TC. Temperature adaptation of synonymous codon usage in different functional categories of genes: a comparative study between homologous genes of *Methanococcus jannaschii* and *Methanococcus maripaludis*. *FEBS Lett* 2006; 580: 3895-3899.
46. Sabbia V, Piovani R, Naya H, Rodriguez-Maseda H, Romero H, Musto H. Trends of amino acid usage in the proteins from the human genome. *J Biomol Struct Dyn* 2007; 25: 55-59.
47. Peden JF. Analysis of codon usage. PhD diss. University of Nottingham 2000.
48. Vesth T, Lagesen K, Acar O, Ussery D. CMG-biotools, a free workbench for basic comparative microbial genomics. *PloS One* 2013; 8: e60120.
49. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986; 3: 418-426.
50. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007; 24: 1586-1591.
51. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 2000; 17: 32-43.
52. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* 1986; 24: 28-38.
53. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng* 1990; 4: 155-161.
54. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003; 31: 3784-3788.
55. Frishman D, Argos P. Seventy-five percent accuracy in protein secondary structure prediction. *Proteins* 1997; 27: 329-335.
56. Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, Liolios K, Pagani I, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* 2012; 40: D123-129.
57. Lynn DJ, Singer GA, Hickey DA. Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Res* 2002; 30: 4272-4277.
58. Paul S, Bag SK, Das S, Harvill ET, Dutta C. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol* 2008; 9: R70.
59. Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987; 15: 1281-1295.
60. Puigbo P, Bravo IG, Garcia-Vallve S. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 2008; 9: 65.
61. Liu WQ, Zhang J, Zhang YQ, Zhou JH, Chen HT, Ma LN, Ding YZ, Liu Y. Compare the differences of synonymous codon usage between the two species within cardiovirus. *Virology* 2011; 8: 325.
62. Nishida H, Beppu T, Ueda K. Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae. *Genomics* 2011; 98: 370-375.
63. De Fariis ST, Bonato MC. Preferred codons and amino acid couples in hyperthermophiles. *Genome Biol* 2002; 3: PREPRINT0006.
64. Sghaier H, Thorvaldsen S, Saied NM. There are more small amino acids and fewer aromatic rings in proteins of ionizing radiation-resistant bacteria. *Annals of Microbiol* 2013; 63: 1483-1491.

65. Kitamura A, Nishimoto M, Sengoku T, Shibata R, Jager G, Bjork GR, Grosjean H, Yokoyama S, Bessho Y. Characterization and structure of the *Aquifex aeolicus* protein DUF752: a bacterial tRNA-methyltransferase (MnmC2) functioning without the usually fused oxidase domain (MnmC1). *J Biol Chem* 2012; 287: 43950-43960.
66. Gentles AJ, Karlin S. Genome-scale compositional comparisons in eukaryotes. *Genome Res* 2001; 11: 540-546.
67. Garcia-Valve S, Guzman E, Montero MA, Romeu A. HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res* 2003; 31: 187-189.
68. Nishida H. Evolution of genome base composition and genome size in bacteria. *Front Microbiol* 2012; 3: 420.
69. Liu Q, Feng Y, Zhao Xa, Dong H, Xue Q. Synonymous codon usage bias in *Oryza sativa*. *Plant Sci* 2004; 167: 101-105.
70. Lu H, Zhao WM, Zheng Y, Wang H, Qi M, Yu XP. Analysis of synonymous codon usage bias in *Chlamydia*. *Acta Biochim Biophys Sin (Shanghai)* 2005; 37: 1-10.
71. 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.
72. Sueoka N, Kawanishi Y. DNA G+C content of the third codon position and codon usage biases of human genes. *Gene* 2000; 261: 53-62.
73. Wan XF, Xu D, Kleinhofs A, Zhou J. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol* 2004; 4: 19.
74. Wu H, Zhang Z, Hu S, Yu J. On the molecular mechanism of GC content variation among eubacterial genomes. *Biol Direct* 2012; 7: 2.
75. Kuo CH, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. *Genome Res* 2009; 19: 1450-1454.
76. Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. *Trends Genet* 2001; 17: 589-596.
77. Sabath N, Ferrada E, Barve A, Wagner A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol Evol* 2013; 5: 966-977.
78. Sakharkar KR, Sakharkar MK, Verma C, Chow VT. Comparative study of overlapping genes in bacteria, with special reference to *Rickettsia prowazekii* and *Rickettsia conorii*. *Int J Syst Evol Microbiol* 2005; 55: 1205-1209.
79. Misra HS, Rajpurohit YS, Kota S. Physiological and molecular basis of extreme radioresistance in *Deinococcus radiodurans*. *Curr Sci (Bangalore)* 2013; 104: 194-205.
80. Krisko A, Smole Z, Debret G, Nikolic N, Radman M. Unstructured hydrophilic sequences in prokaryotic proteomes correlate with dehydration tolerance and host association. *J Mol Biol* 2010; 402: 775-782.
81. Venkateswaran A, McFarlan SC, Ghosal D, Minton KW, Vasilenko A, Makarova K, Wackett LP, Daly MJ. Physiologic determinants of radiation resistance in *Deinococcus radiodurans*. *Appl Environ Microbiol* 2000; 66: 2620-2626.
82. Gromiha MM, Thomas S, Santhosh C. Role of cation-pi interactions to the stability of thermophilic proteins. *Prep Biochem Biotech* 2002; 32: 355-362.