Available online at www.sciencedirect.com
ScienceDirect
Genomics 90 (2007) 103–109
GENOMICS
www.elsevier.com/locate/ygeno

Genomic survey of sequence features for ultraviolet tolerance in haloarchaea (family Halobacteriaceae)

Peng Zhou [a], Ji Wen [a], Aharon Oren [b], Ming Chen [c], Min Wu [a,*]

[a] Department of Microbiology, College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[b] Institute of Life Sciences and the Moshe Shilo Minerva Center for Marine Biogeochemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
[c] Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China

Received 18 December 2006; accepted 27 March 2007
Available online 11 May 2007

Abstract

We have investigated the strategy of *Halobacterium* sp. NRC-1 and other members of the family Halobacteriaceae to survive ultraviolet (UV) irradiation, based on an integrated analysis of various genomic and proteomic features such as dinucleotide composition and distribution of tetranucleotides in the genome and amino acid composition of the proteins. The low dipyrimidine content may help *Halobacterium* reduce formation of photoproducts in its genome. The usage of residues susceptible to reactive oxygen species attack is reduced significantly in *Halobacterium*, which helps the organism to minimize protein damage. We then correlated the expression of the *zim* gene with the genomic structure to reexamine the importance of the putative mismatch repair pathway proposed previously. Our results showed that *Halobacterium* sp. NRC-1 and other haloarchaea (*Haloarcula marismortui*, *Haloquadratum walsbyi*) have optimized their genomic and proteomic structures to reduce damage induced by UV irradiation, often present at high levels in habitats where these organisms thrive.
© 2007 Elsevier Inc. All rights reserved.

Keywords: *Halobacterium*; UV tolerance; Genomic structure; Dipyrimidine composition; CTAG pattern; Amino acid composition

Halophilic archaea (haloarchaea) of the family Halobacteriaceae (genera *Halobacterium*, *Haloferax*, *Haloarcula*, *Halorubrum*, and others) live in hypersaline environments such as salt lakes and salterns, typically containing over 4.5 M NaCl, where they are exposed to high solar radiation and to oxidative stress. The adaptive mechanisms enabling *Halobacterium* and other members of the family to survive in their habitats have been extensively investigated [1]. Their radiation tolerance has been intensively studied from the 1980s onward [2–7]. In recent years, *Halobacterium* sp. NRC-1, the first member of the family whose genome has been completely sequenced, has become a popular model organism for such studies.

The survival rate of *Halobacterium* sp. NRC-1 is near 100% following irradiation with shortwave ultraviolet light (UV-C) doses up to 110 J/m^2, and UV-C lesions are very efficiently photoreactivated by an active cyclobutane pyrimidine dimers (CPD) photolyase, *phr*2 [8]. In addition to photoreactivation, homologs of the light-independent bacterial nucleotide excision repair (NER) genes *uvrA*, *uvrB*, and *uvrC* have been shown to be required for the repair of UV damage in the absence of photoreactivating light [9]. Eukaryotic NER orthologs are also present in *Halobacterium* sp. NRC-1 [10]. Analysis of transcriptional profiling and protein abundance changes revealed that RadA1-mediated homologous recombination plays an important role in the cellular response to 30–70 J/m^2 of UV-C or 2.5 kGy of gamma rays [8,11].

Since there are so many genes and pathways related to DNA repair in *Halobacterium* sp. NRC-1, and previous studies showed the UV photoprotection afforded by membrane pigments, the question whether *Halobacterium* may possess some direct mechanism to reduce the appearance of lesions, in addition to their repair, arises. If it does, then whether such a mechanism may be common to other halophilic archaea living in similar habitats should be examined.

To answer these questions, our investigation focused on those hot spots in DNA where UV lesions commonly occur. UV-C induces two main types of mutagenic lesions in DNA: CPD formed between adjacent thymidine or cytosine residues,

* Corresponding author. Fax: +86 571 88206134 8000.
E-mail address: wumin@zju.edu.cn (M. Wu).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.03.015

and pyrimidine (6–4) pyrimidone (6–4) photoproducts formed between adjacent pyrimidine residues, mostly between T-C and C-C residues [12].

In this study, we surveyed the dipyrimidine composition in the genomes of *Halobacterium* sp. NRC-1 and other extreme halophilic archaea, using the intestinal bacterium *Escherichia coli* K12, which is not exposed to sunlight in its natural habitat, for comparison. We found that the low occurrence of dipyrimidine in the genomes of haloarchaea reduces the probability of the occurrence of photoproducts.

In addition to lesions in DNA, UV irradiation can also cause photooxidative damage in proteins. Several oxidative damage repair enzymes and proteases for eliminating damaged proteins have been identified in *Halobacterium* sp. NRC-1, including the superoxide dismutases Sod1 and Sod2 and the putative serine-protease Lon, hypothesized to play a role in degrading photooxidized proteins [8,13,14]. Inspired by previous studies that showed that the high density of acidic residues on the surfaces of almost all its proteins helps to stabilize their structures and functions in high-salinity habitats [15], we here address the question whether haloarchaea may have evolved some bias of amino acid usage to better resist the oxidative stress and whether the bias exists as a common phenomenon among organisms threatened by oxidative pressure. We therefore examined predicted amino acid compositions in the proteins of the chosen organisms to assess the usage of amino acid residues prone to reactive oxygen species (ROS) attack [16].

In addition to the analysis of dipyrimidine and amino acid composition, we investigated the previous transcriptional data and the GATC and CTAG distribution pattern. A study in yeast suggested that most UV inducible genes do not contribute to survival following UV irradiation [17], and most genes involved in surviving UV damage, including most NER genes, are not UV inducible. The same is true for human cells [18]. Therefore the profile of post-UV-irradiation gene expression in *Halobacterium* should be interpreted with caution, and inferences about the extent of gene involvement should be supported by physiological and functional studies [19]. With regard to the putative d(CTAG) methylation-directed mismatch repair (MMR) hypothesized on the basis of transcriptome analysis, we examined the expression of the *zim* gene, the homolog of CTAG-specific methylase, which did not show much significance, according to the general selection rule used in previous study in silico [8]. Moreover, we surveyed the distribution of the sequence CTAG in the genomes of halophilic archaea involved in the putative d(CTAG) methylation-directed MMR mechanism [8].

By mining genomic and proteomic features in haloarchaea we here show how these organisms reduce the occurrence of lesions induced by UV irradiation in their habitats.

Results

Dipyrimidine composition

We have analyzed the dipyrimidine frequency in the largest replicon from each organism listed in Table 1. The biases of dipyrimidine in the haloarchaeal genomes were similar but distinct from those in *Escherichia coli* as shown in Fig. 1. The frequency of TT observed in the extreme halophilic archaea is particularly low. For example, in the chromosome of *Halobacterium* sp. NRC-1, the bias of TT is −7.7%, that for CT is −21.3%, and that for CC is −22.3%, while in *E. coli* K12 the bias of TT is 20.6%. The frequencies of CT and CC are less than expected in all four chromosomes. We further observed a much higher bias of TC in the extremely halophilic archaea and a far lower value in *E. coli* K12.

To reveal the possible reason for the increase of TC, we further examined the 16 codons involving TC and its complementary code, GA. Among the 16 codons, 4 code for Ser, 2 for Asp, 2 for Glu, 2 for Arg, 1 for Phe, 1 for Leu, 1 for Ile, 1 for Val, and 1 for Gly, and the remaining one is a stop codon. By comparison with the codon bias plot [15], we found that the increase in TC is caused by the overrepresentation of acidic residues, Asp and Glu, coded by GAU, GAC, GAA, and GAG. Thus, it is necessary in the survey of the dipyrimidine frequency in haloarchaea and its impact on UV resistance to take into account the contribution of the overrepresentation of acidic residues to the number of TCs. Therefore we performed regression analysis to examine the linear correlation between the biases of TC and the frequency of acidic residues. By regression analysis of haloarchaeal replicons, we obtained a

Table 1
Replicons used for genomic dipyrimidine composition and amino acid composition analysis

Species	Replicons	RefSeq	Size (bp)	Version
Deinococcus radiodurans R1	chromosome 1	NC_001263	2648638	3-Dec-05
	chromosome 2	NC_001264	412348	3-Dec-05
	plasmid CP1	NC_000959	45704	29-Oct-04
	plasmid MP1	NC_000958	177466	21-Jun-06
Halobacterium sp. NRC-1	chromosome	NC_002607	2014239	2-Dec-05
	Plasmid pNRC100	NC_001869	191346	2-Dec-05
	plasmid pNRC200	NC_002608	365425	2-Dec-05
Har. marismortui ATCC 43049	chromosome I	NC_006396	3131724	18-Jan-06
	chromosome II	NC_006397	288050	30-Mar-06
	plasmid pNG100	NC_006389	33303	30-Mar-06
	plasmid pNG200	NC_006390	33452	30-Mar-06
	plasmid pNG300	NC_006391	39521	30-Mar-06
	plasmid pNG400	NC_006392	50060	30-Mar-06
	plasmid pNG500	NC_006393	132678	30-Mar-06
	plasmid pNG600	NC_006394	155300	3-Dec-05
	plasmid pNG700	NC_006395	410554	3-Dec-05
Hqr. walsbyi DSM 16790	chromosome	NC_008212	3132494	24-Aug-06
	plasmid PL47	NC_008213	46867	24-Aug-06
Synechococcus elongatus PCC 7942	chromosome	NC_007604	2695903	15-Jun-06
	plasmid 1	NC_007595	46366	30-Mar-06
E. coli K12	chromosome	NC_000913	4639675	22-Sep-06

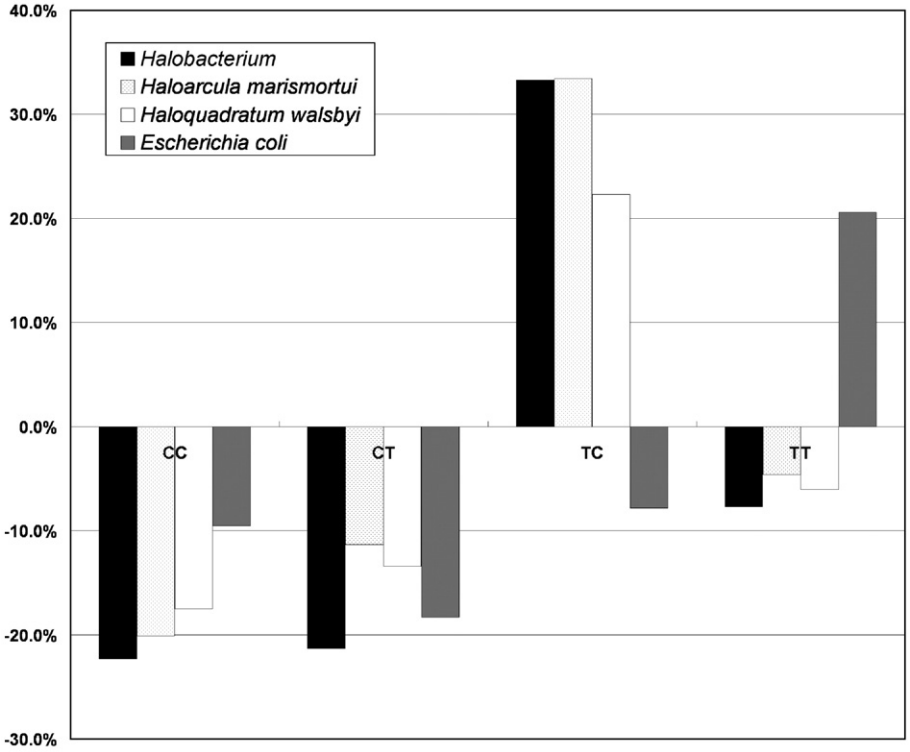

Fig. 1. Dipyrimidine composition of the largest replicon of the analyzed genomes. Each pattern of dipyrimidine is shown on the *x*-axis and the bars show the corresponding bias.

trend line of $y=0.5577x-0.3166$ with correlation coefficient R equal to 0.9136, evaluated by the two-tailed t test ($t=4.6504$, $t_{0.05,\,4}=2.7764$) (Fig. 2). We examined whether the bias of TC is lower than zero when the bias of the frequency of acidic residues approaches zero, as reflected by the intercept. We performed the one-tailed t test and found a p value of 0.0409,

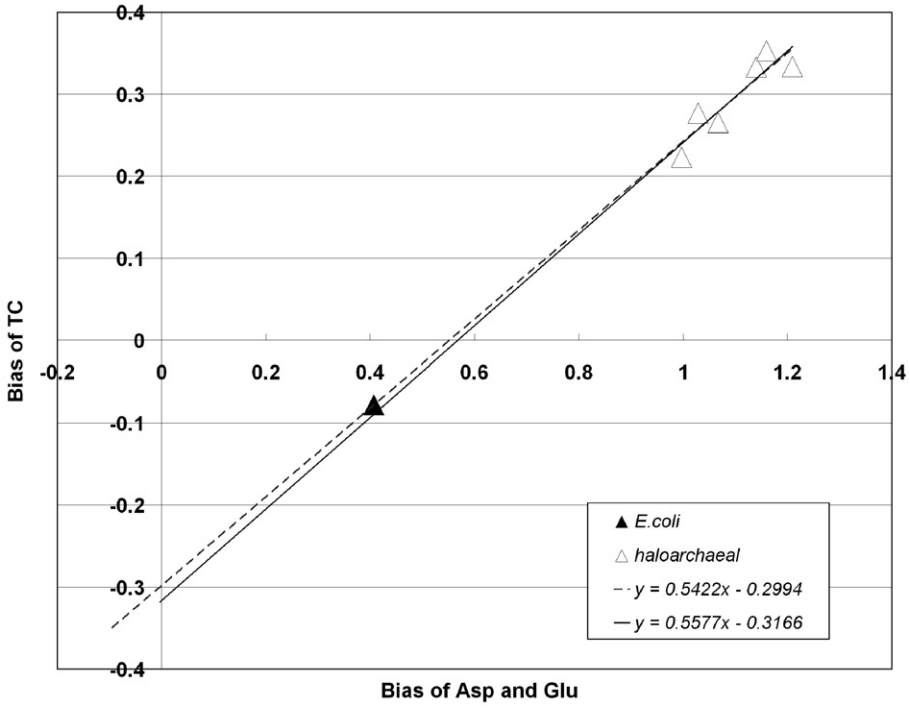

Fig. 2. Bias of TC versus bias of Asp and Glu for haloarchaeal replicons (empty triangles) including *Halobacterium* sp. NRC-1, *Har. marismortui* ATCC 43049, *Hqr. walsbyi* DSM 16790, and *E. coli* K12 (filled triangle). The solid line represents the regression trendline of the bias of TC versus the bias of Asp and Glu in haloarchaeal replicons. The dashed line represents the regression trendline of the bias of TC against the bias of Asp and Glu in *E. coli* K12 and the haloarchaeal replicons.

the intercept being in the interval $(-\infty, -0.0247)$ with 95% confidence, which suggested that, when the observed value of acidic amino acids is equal to expected (bias of Asp and Glu=0), the bias of TC would be in the interval $(-\infty, -0.0247)$ with a probability of 95%, and the observed value of TC is still lower than expected.

When the genome of *E. coli* K12 was included in the regression analysis, we obtained a second trend line of $y=0.5422x-0.2994$ with correlation coefficient R equal to 0.9921, the p value of the intercept being 0.0002, with 95% confidence interval $(-0.3809, -0.2179)$.

From the two trend lines in Fig. 2, obviously the bias of TC gradually decreases as the acidic amino acid composition approaches the expected values. The intercept of line 1, −0.2994, is larger than that of line 2, −0.3166, showing that haloarchaea would have a lower content of TC than *E. coli* K12 if the effect of the excess acidic amino acid is eliminated.

Amino acid composition

A large excess of acidic amino acids in proteins is a common phenomenon among extremely halophilic archaea [15]. Interestingly, we also observed that those residues prone to ROS attack, namely Phe, Arg, His, Lys, Met, Pro, Trp, Tyr, and Cys, displayed lower frequencies in the proteins of the haloarchaea than expected (Fig. 3). Bias of Cys is the lowest among the 20 amino acids. In *Halobacterium* sp. NRC-1, the bias of Cys is −79.5%. Trp was present −54.9%, Lys −51.5%, His −39.3%, Met −29.8%, Tyr −29.7, Arg −24.4%, Pro −24.3%, and Phe −15.0%. Compared with the proteins in *E. coli* K12, Lys, Met, and Phe were significantly decreased in haloarchaea. The amino acids susceptible to ROS were also less used than expected in *Synechococcus elongatus* PCC 7942 and in *Deinococcus radiodurans* R1, organisms threatened by oxygen radicals in vivo caused by desiccation or metabolism [20,21]. As the open reading frames (ORFs) predicted from the genome sequences were not always consistent with the real profiles of the translated proteins, we selected the proteins identified from studies of *Halobacterium* sp. NRC-1 [13,14,22] and analyzed the amino acid composition of the 975 proteins selected, with altogether 315,175 amino acids (See Supplementary Table). The result showed that amino acids prone to ROS attack in the 975 proteins were less used than those in all the ORFs, Lys being the only exception (Fig. 3). In addition to the amino acids susceptible to ROS, Ser was also less used among the proteins. Asn and Gln were both less used in halophilic archaea. In *Hqr. walsbyi* there is a sharp interesting increase of usage of Ile contrasting to the decrease in the other halophilic archaea.

CTAG distribution pattern

We investigated the CTAG distribution pattern involved in the putative d(CTAG) MMR in *Halobacterium* sp. NRC-1 and compared it with the GATC distribution pattern directing d(GATC) MMR in *E. coli*. There were 991 sites of CTAG found across the genome of *Halobacterium* sp. NRC-1. The average fragment length was 2594 bp, with 249 fragments longer than 3000 bp. The numbers of fragments versus their corresponding lengths are shown in Fig. 4. The three longest distances observed were 29,975, 28,884, and 26,668 bp. Genes encoded by the three fragments are involved in transcription, translation, and other important metabolic processes. For example, genes

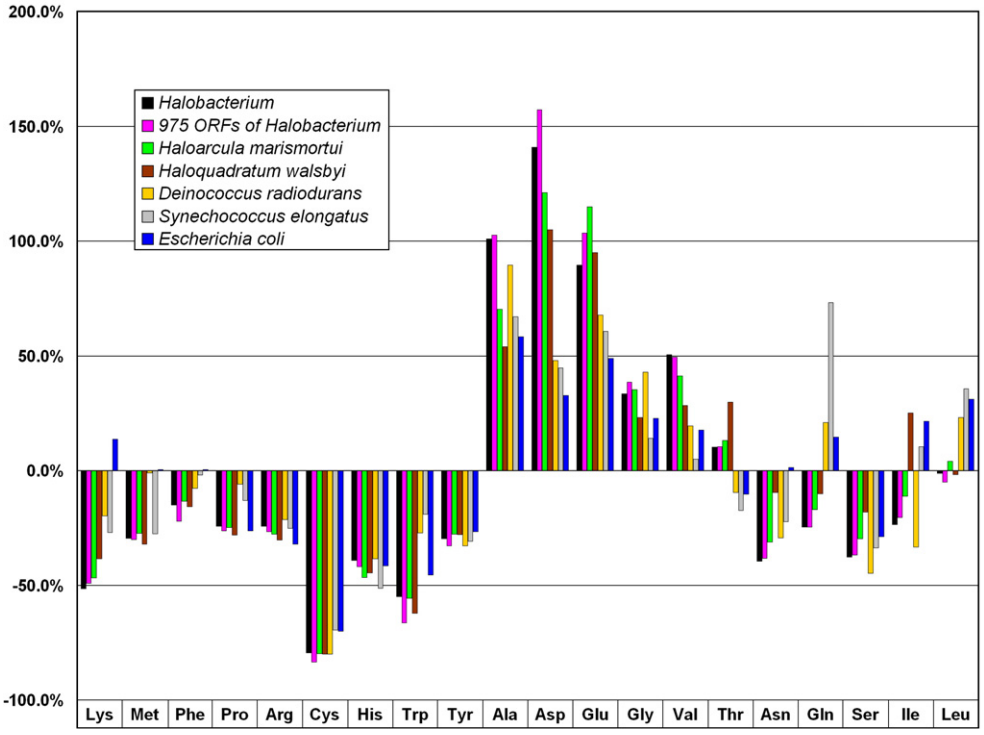

Fig. 3. Biases of 20 amino acids in the proteomes chosen. Over- and underrepresented amino acid usage is denoted as positive and negative values, respectively.

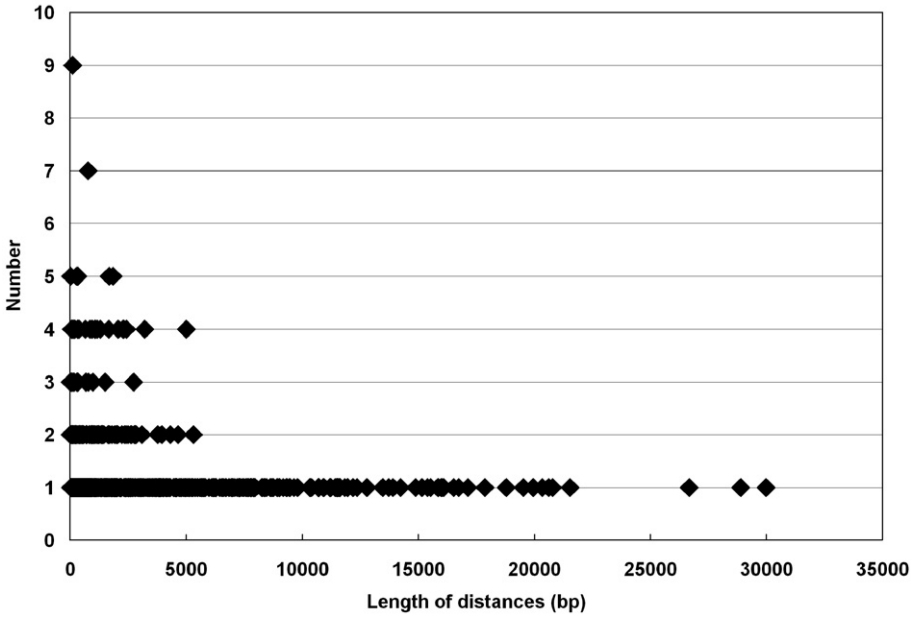

Fig. 4. Distances between every two CTAG sequences in the range of the *Halobacterium* strain NRC-1 genome. The numbers of fragments with the same length are summed and shown as black diamonds.

for many single-copy ribosomal proteins are located within the three fragments, more than 1500 bp removed from either CTAG (Table 2). For comparison, the total number of GATC sites in *E. coli* is 18,711 and the average fragment length between two successive sites is 248 bp. Only 7 fragments are longer than 3000 bp, with the longest being 4836 bp.

Discussion

Halophilic archaea of the family Halobacteriaceae, which inhabit shallow lakes and therefore are often exposed to harmful levels of UV radiation, have evolved essential mechanisms of survival, including strategies of avoidance, protection, and repair. Studies on *Halobacterium* sp. NRC-1 showed that in this organism avoidance includes negative phototaxis away from the source of UV radiation, and the sensors involved have been identified [10,23]. This mechanism is not present in all species: *Hfx. volcanii* and *Hqr. walsbyi* are nonmotile, and therefore they lack a motility-based escape strategy. Until now, few attempts to interpret the UV tolerance of haloarchaea from the genomic and protein structures point of view have been made. In this study we examined these mechanisms from the genomic and protein structures perspectives.

Table 2
Ribosomal proteins encoded by the three longest CTAG fragments in the *Halobacterium* strain NRC-1 genome

Gene name	Position in chromosome	Function
rps8e	1244984 << 1245355	30S ribosomal protein S8E
rpl3p	1259033 >> 1260037	50S ribosomal protein L13P
rpl4e	1260044 >> 1260793	50S ribosomal protein L4E
rpl23p	1260793 >> 1261044	50S ribosomal protein L23P
rpl2p	1261050 >> 1261769	50S ribosomal protein L2P
rps19p	1261776 >> 1262195	30S ribosomal protein S19P
rpl22p	1262202 >> 1262669	50S ribosomal protein L22P
rps3p	1262672 >> 1263577	30S ribosomal protein S3P
rpl29p	1263580 >> 1263792	50S ribosomal protein L29P
rps17p	1264166 >> 1264492	30S ribosomal protein S17P
rpl14p	1264495 >> 1264890	50S ribosomal protein L14P
rpl24p	1264898 >> 1265251	50S ribosomal protein L24P
rps4e	1265251 >> 1265949	30S ribosomal protein S4E
rpl5p	1265949 >> 1266473	50S ribosomal protein L5P
rps14p	1266473 >> 1266628	30S ribosomal protein S14P
rps27ae	1508438 << 1508569	30S ribosomal protein S27E
rps24e	1508574 << 1508879	30S ribosomal protein S24E
rpl40e	1526849 >> 1526989	50S ribosomal protein L40E

The mechanisms of UV tolerance existing in the genomic and proteomic structures lower the occurrence of potential lesions even before the repair process, but the efficacy of these mechanisms is limited. Once lesions occur, the repair-related genes and pathways will play important roles in eliminating the damaged cell components, such as pyrimidine dimers and single- or double-stranded breaks in the DNA and photooxidized proteins [8,9,11,19,23–26]. But while interpreting the repair mechanisms, we should examine the relevance of pathways such as MMR by analysis of the structure of the genome. In typical d(GATC)-directed MMR, mismatch to d(GATC) site distance was usually shorter than 1 kb [27], consistent with our results. Beyond distance, the effect of mismatch correction was almost unnoticeable. In some instances, stretches of DNA as long as 1500 nucleotides may be replaced by long patch excision repair [28]. In comparison, in *Halobacterium* sp. NRC-1 we found 249 fragments longer than 3000 bp. Thus, a number of questions could be posed. If the d(CTAG) methylation-directed MMR does exist, how could it repair distantly located errors, such as are expected to occur within a stretch of 29,975 bp between two distantly separated CTAGs, the longest one shown in Fig. 4? Could it be possible that the putative CTAG MMR

system is efficient and capable of repairing longer stretches of DNA and is there another MMR pathway, for example the typical d(GATC) MMR? The pathway should be reconsidered in further study.

While we found that the haloarchaea adopt a low dipyrimidine composition to reduce formation of possible photoproducts in their genomes, we also interpreted the increase in the number of TCs as being caused by the demand for the corresponding acidic residues, which are essential in maintaining the function of proteins in the high-salt environment [15]. We note that the acidic amino acids are not among the nine amino acids easily attacked by ROS [16]. The haloarchaea thus simultaneously solve the problem of the need for acidic amino acids to enable protein conformation and function and the need to stabilize both their genomes and their proteins against UV damage. The usage of Lys residues is higher than expected in *E. coli*, but the other eight amino acids susceptible to ROS were equally or less used than expected. As the production of ROS during cell metabolism cannot be avoided, many organisms have evolved a low content of residues susceptible to ROS, together with the genes encoding ROS scavenging mechanisms.

Some extremophiles have to adapt to resist multiple environmental stress factors. In the case of the halophilic archaea these factors may include, in addition to salinity, extreme pH, temperature, UV radiation levels, and heavy metal concentrations. The result of the selection process required to withstand such additional stress factors is reflected in the genomic and protein structures of the organisms inhabiting such environments. The methods of correlation analysis used in this study will undoubtedly be helpful toward the elucidated of how such extremophiles adapt their structures to their stressful habitats.

Finally, with the rapid increase of the number of complete genomes of extremophiles [29], genomic and proteomic signature analysis will lead to an understanding of how the extremophiles adapt to their habitats.

Materials and methods

Genome sequences

Genome sequences of *Halobacterium* sp. NRC-1 (ATCC 700922) [10], *Har. marismortui* (ATCC 43049^T) [30], and *Hqr. walsbyi* (DSM 16790^T) [31], all organisms having similar salinity, pH, and temperature optima, and the intestinal bacterium *E. coli* K12 [32] were derived from GenBank. Details of the replicons are shown in Table 1.

Dipyrimidine composition

The genomic profile of dipyrimidine consists of the array $\{\rho^*_{XY}=f^*_{XY}/f^*_X f^*_Y\}$, where f^*_X denotes the frequency of the monopyrimidine X and f^*_{XY} the frequency of the dipyrimidine XY, both computed from the double circular strands of each replicon. These dipyrimidine relative abundance values $\{\rho^*_{XY}\}$ minus 1 (termed dipyrimidine biases) effectively assess differences between the observed dipyrimidine frequencies and those expected from random associations of the component monopyrimidine frequencies [33,34].

Regression analysis was performed using Microsoft Excel [35] to calculate the linear correlations between biases of TC and acidic amino acids in six replicons in the four species. The six replicons include the chromosomes of three species and the two largest plasmids longer than 280,000 bp which are approximately equal in size to the shortest chromosome, chromosome II in *Har. marismortui*. The total biases of acidic amino acids of six replicons were calculated as above, using the values of observation and expectation calculated in the next part of amino acid composition. The correlation coefficient R was evaluated by the two-tailed t test according to

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}},$$

where r and n mean correlation coefficient and number of the samples, respectively.

Amino acid composition

All ORFs were separately extracted from the chosen complete genomes and entered into the DAMBE Software (Version 4.2.13) [36]. Then the expected amino acid frequencies were calculated by using the "*Seq.Analysis | Amino acid frequency*" option in the software. The calculation of the expectation frequency was based on the fact that the frequency of amino acids is positively correlated with the number of codons per amino acid. The output tests this prediction by calculating the Pearson correlation coefficient and doing a simple linear regression of frequency of amino acid on number of codons per amino acid. Because a positive correlation between the two has already been predicted, the test is one-tailed [37]. Similar to dipyrimidine biases, the bias of each amino acid was calculated to detect over- and underrepresented amino acid usage denoted as positive and negative values in Fig. 3. These calculations were performed for 975 proteins expressed, composed of 317,294 amino acids, by integrating proteomics analysis of *Halobacterium* sp. NRC-1 [13,14,22] (See Supplementary Table). The composition of the 975 ORFs was calculated as described above.

CTAG pattern

The length of each fragment between successive CTAG sites was calculated by Restriction Digest Tool at http://cmr.tigr.org and Microsoft Excel.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program, Grant No. 2004CB719604-3) and in part by the National Science Foundation of China (Grant No. 30670048, No. 30370029, No. 30500106). We are grateful to Nitin S. Baliga (Institute for Systems Biology, Seattle, WA, USA), and Wailap Victor Ng (Notional YangMing University, Taipei, China) for suggestions in collecting data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.03.015.

References

[1] S. DasSarma, B.R. Berquist, J.A. Coker, P. DasSarma, J.A. Müller, Post-genomics of the model haloarchaeon *Halobacterium* sp. NRC-1, Saline Systems 2 (2006) 3.
[2] P.S. Fitt, N. Sharma, G. Castellanos, A comparison of liquid-holding recovery and photoreactivation in halophilic and non-halophilic bacteria, Biochim. Biophys. Acta 739 (1983) 73–78.
[3] N. Sharma, D. Hepburn, P.S. Fitt, Photoreactivation in pigmented and non-pigmented extreme halophiles, Biochim. Biophys. Acta 799 (1984) 135–142.
[4] P.S. Fitt, N. Sharma, The fate of thymine-containing dimers in ultraviolet-

irradiated *Halobacterium cutirubrum*, Biochim. Biophys. Acta 910 (1987) 103–110.
[5] S. McCready, The repair of ultraviolet light-induced DNA damage in the halophilic archaebacteria, *Halobacterium cutirubrum, Halobacterium halobium* and *Haloferax volcanii*, Mutat. Res. 364 (1996) 25–32.
[6] E.L. Martin, R.L. Reinhardt, L.L. Baum, M.R. Becker, J.J. Shaffer, T.A. Kokjohn, The effects of ultraviolet radiation on the moderate halophile *Halomonas elongata* and the extreme halophile *Halobacterium salinarum*, Can. J. Microbiol. 46 (2000) 180–187.
[7] A.P.M. Eker, L. Formenoy, L.E.A. Wit, Photoreactivation in the extreme halophilic archaebacterium *Halobacterium cutirubrum*, Photochem. Photobiol. 53 (1991) 643–651.
[8] N.S. Baliga, S.J. Bjork, R. Bonneau, M. Pan, C. Iloanusi, M.C. Kottemann, L. Hood, J. DiRuggiero, Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium* NRC-1, Genome Res. 14 (2004) 1025–1035.
[9] D.J. Crowley, I. Boubriak, B.R. Berquist, M. Clark, E. Richard, L. Sullivan, S. DasSarma, S. McCready, The *uvrA, uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1, Saline Systems 2 (2006) 11.
[10] W.V. Ng, S.P. Kennedy, G.G. Mahairas, B. Berquist, M. Pan, H.D. Shukla, S.R. Lasky, N.S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T.A. Dahl, R. Welti, Y.A. Goo, B. Leithauser, K. Keller, R. Cruz, M.J. Danson, D.W. Hough, D.G. Maddocks, P.E. Jablonski, M.P. Krebs, C.M. Angevine, H. Dale, T.A. Isenbarger, R.F. Peck, M. Pohlschröder, J.L. Spudich, K.W. Jung, M. Alam, T. Freitas, S. Hou, C.J. Daniels, P.P. Dennis, A.D. Omer, H. Ebhardt, T.M. Lowe, P. Liang, M. Riley, L. Hood, S. DasSarma, Genome sequence of *Halobacterium* species NRC-1, Proc. Natl. Acad. Sci. USA 97 (2000) 12176–12181.
[11] K. Whitehead, A. Kish, M. Pan, A. Kaur, D.J. Reiss, N. King, L. Hohmann, J. DiRuggiero, N.S. Baliga, An integrated systems approach for understanding cellular responses to gamma radiation, Mol. Syst. Biol. 2 (2006) 47.
[12] E. Friedberg, G.C. Walker, W. Siede, DNA repair and mutagenesis, ASM Press, Washington, D.C., 1995
[13] Y.A. Goo, E.C. Yi, N.S. Baliga, W.A. Tao, M. Pan, R. Aebersold, D.R. Goodlett, L. Hood, W.V. Ng, Proteomic analysis of an extreme halophilic archaeon, *Halobacterium* sp. NRC-1, Mol. Cell. Proteomics 2 (2003) 506–524.
[14] R.R. Gan, E.C. Yi, Y. Chiu, H. Lee, Y.C. Kao, T.H. Wu, R. Aebersold, D.R. Goodlett, W.V. Ng, Proteome analysis of *Halobacterium* sp. NRC-1 facilitated by the biomodule analysis tool BMSorter, Mol. Cell. Proteomics 5 (2006) 987–997.
[15] S.P. Kennedy, W.V. Ng, S.L. Salzberg, L. Hood, S. DasSarma, Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence, Genome Res. 11 (2001) 1641–1650.
[16] R.L. Zheng, Z.Y. Huang, Basics of Free Radical Medicine and Agriculture, Higher Education Press Beijing and Springer-Verlag, Heidelberg, 2001, p. 12.
[17] G.W. Birrell, J.A. Brown, H.I. Wu, G. Giaever, A.M. Chu, R.W. Davis, J.M. Brown, Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents, Proc. Natl. Acad. Sci. USA 99 (2002) 8778–8783.
[18] K.E. Rieger, G. Chu, Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells, Nucleic Acids Res. 32 (2004) 4786–4803.
[19] S. McCready, J.A. Müller, I. Boubriak, B.R. Berquist, W.L. Ng, S. DasSarma, UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium* sp. NRC-1, Saline Systems 1 (2005) 3.
[20] K.P. Michel, E.K. Pistorius, S.S. Golden, Unusual regulatory elements for iron deficiency induction of the *idiA* gene of *Synechococcus elongatus* PCC 7942, J. Bacteriol. 183 (2001) 5015–5024.
[21] O. White, J.A. Eisen, J.F. Heidelberg, E.K. Hickey, J.D. Peterson, R.J. Dodson, D.H. Haft, M.L. Gwinn, W.C. Nelson, D.L. Richardson, K.S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J.J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K.S. Makarova, L. Aravind, M.J. Daly, K.W. Minton, R.D. Fleischmann, K.A. Ketchum, K.E. Nelson, S. Salzberg, H.O. Smith, J.C. Venter, C.M. Fraser, Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1, Science 286 (1999) 1571–1577.
[22] H.D. Shukla, Proteomic analysis of acidic chaperones, and stress proteins in extreme halophile *Halobacterium* NRC-1: a comparative proteomic approach to study heat shock response, Proteome Sci. 4 (2006) 6.
[23] N.S. Baliga, M. Pan, Y.A. Goo, E.C. Yi, D.R. Goodlett, K. Dimitrov, P. Shannon, R. Aebersold, W.V. Ng, L. Hood, Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach, Proc. Natl. Acad. Sci. USA 99 (2002) 14913–14918.
[24] J.A. Müller, S. DasSarma, Genomic analysis of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors, J. Bacteriol. 187 (2005) 1659–1667.
[25] G. Wang, S.P. Kennedy, S. Fasiludeen, C. Rensing, S. DasSarma, Arsenic resistance in *Halobacterium* sp. strain NRC-1 examined by using an improved gene knockout system, J. Bacteriol. 186 (2004) 3187–3194.
[26] A. Kaur, M. Pan, M. Meislin, M.T. Facciotti, R. El-Gewely, N.S. Baliga, A systems view of haloarchaeal strategies to withstand stress from transition metals, Genome Res. 16 (2006) 841–854.
[27] R. Bruni, D. Martin, J. Jiricny, d(GATC) sequences influence *Escherichia coli* mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch, Nucleic Acids Res. 16 (1988) 4875–4890.
[28] P.K. Cooper, Characterization of long patch excision repair of DNA in ultraviolet-irradiated *Escherichia coli*: an inducible function under rec-lex control, Mol. Gen. Genet. 185 (1982) 189–197.
[29] R. Pinard, A. de Winter, G.J. Sarkis, M.B. Gerstein, K.R. Tartaro, R.N. Plant, M. Egholm, J.M. Rothberg, J.H. Leamon, Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing, BMC Genomics 7 (2006) 216.
[30] N.S. Baliga, R. Bonneau, M.T. Facciotti, M. Pan, G. Glusman, E.W. Deutsch, P. Shannon, Y. Chiu, R.S. Weng, R.R. Gan, P. Hung, S.V. Date, E. Marcotte, L. Hood, W.V. Ng, Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea, Genome Res. 14 (2004) 2221–2234.
[31] H. Bolhuis, P. Palm, A. Wende, M. Falb, M. Rampp, F. Rodríguez-Valera, F. Pfeiffer, D. Oesterhelt, The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity, BMC Genomics 7 (2006) 169.
[32] F.R. Blattner, G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, The complete genome sequence of *Escherichia coli* K-12, Science 277 (1997) 1453–1474.
[33] S. Karlin, C. Burge, Dinucleotide relative abundance extremes: a genomic signature, Trends Genet. 11 (1995) 283–290.
[34] A. Campbell, J. Mrazek, S. Karlin, Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA, Proc. Natl. Acad. Sci. USA 96 (1999) 9184–9189.
[35] W.Z. Wang, Applications of Excel in statistical analysis, China Railway Publishing House, Beijing, 2002.
[36] X. Xia, Z. Xie, DAMBE: software package for data analysis in molecular biology and evolution, J. Hered. 92 (2001) 371–373.
[37] X. Xia, Data analysis in molecular biology and evolution, Kluwer Academic Publishers, Boston, 2000.