Plasmid 65 (2011) 77–101

Contents lists available at ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Review

The information transfer system of halophilic archaea

Melinda D. Capes [a,b], James A. Coker [a,1], Rachael Gessler [a], Valerie Grinblat-Huse [a,b], Satyajit L. DasSarma [a], Christopher G. Jacob [b], Jong-Myoung Kim [c], Priya DasSarma [a], Shiladitya DasSarma [a,b,*]

[a] Department of Microbiology & Immunology, University of Maryland School of Medicine, Baltimore, MD, USA
[b] Graduate Program in Life Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
[c] Department of Marine Bio-Materials & Aquaculture, PuKyong National University, Pusan, Republic of Korea

ARTICLE INFO

Article history:
Received 27 August 2010
Accepted 15 November 2010
Available online 19 November 2010

Keywords:
Haloarchaea
Comparative genomics
DNA replication
DNA repair
Transcription
Translation

ABSTRACT

Information transfer is fundamental to all life forms. In the third domain of life, the archaea, many of the genes functioning in these processes are similar to their eukaryotic counterparts, including DNA replication and repair, basal transcription, and translation genes, while many transcriptional regulators and the overall genome structure are more bacterial-like. Among halophilic (salt-loving) archaea, the genomes commonly include extrachromosomal elements, many of which are large megaplasmids or minichromosomes. With the sequencing of genomes representing ten different genera of halophilic archaea and the availability of genetic systems in two diverse models, *Halobacterium* sp. NRC-1 and *Haloferax volcanii*, a large number of genes have now been annotated, classified, and studied. Here, we review the comparative genomic, genetic, and biochemical work primarily aimed at the information transfer system of halophilic archaea, highlighting gene conservation and differences in the chromosomes and the large extrachromosomal elements among these organisms.

Contents

* Corresponding author. Address: 710 E Pratt St., Suite 236, Baltimore, MD 21202, USA. Fax: +1 410 234 8896.
E-mail address: sdassarma@som.umaryland.edu (S. DasSarma).
[1] Present address: Department of Biology, University of Alabama at Birmingham, AL, USA.

0147-619X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.plasmid.2010.11.005

1. Introduction

Halophilic archaea (haloarchaea) thrive in extremely saline environments such as the Dead Sea and Great Salt Lake where salt concentrations reach near saturating conditions (DasSarma and DasSarma, 2006; DasSarma et al., 2009). In addition to high salinity, several of these microorganisms are also faced with other extreme conditions, such as *Halorubrum lacusprofundi*, which is adapted to thrive at low temperatures (Franzmann et al., 1988), and *Natronomonas pharaonis*, which is adapted to thrive in alkaline conditions (Soliman and Truper, 1982). The genome sequence of the first sequenced haloarchaeon (*Halobacterium* sp. NRC-1), a typical species isolated from sea salt, revealed that it had significant similarities to eukaryotes with regard to its information transfer system, yet was similar to bacteria with regard to its genome structure, possessing a circular chromosome and two large plasmids (Ng et al., 2000; DasSarma, 2004).

Since the first haloarchaeal genome was sequenced, genomes for eight additional haloarchaeal genera have been published: *Haloarcula marismortui* (Baliga et al., 2004), *N. pharaonis* (Falb et al., 2005), *Haloquadratum walsbyi* (Bolhuis et al., 2006), *Halogeometricum borinquense* (Malfatti et al., 2009), *Halomicrobium mukohataei* (Tindall et al., 2009), *Halorhabdus utahensis* (Bakke et al., 2009), *Haloferax volcanii* (Hartman et al., 2010), and *Haloterrigena turkmenica* (Hartman et al., 2010; Saunders et al., 2010). The completed genome sequence of an additional haloarchaeon, *H. lacusprofundi*, is available through NCBI as is the genome of *Halobacterium* sp. R-1, a species that is similar to but distinct from *Halobacterium* sp. NRC-1 (Ng et al., 2008; Pfeiffer et al., 2008). All of these sequenced haloarchaea are extreme halophiles, growing optimally in NaCl concentrations ranging from 2.5 to 4.6 M.

The sequenced haloarchaeal genomes are mostly composed of relatively small circular chromosomes and usually contain large extrachromosomal elements (megaplasmids), many of which are 100 kb or larger. The extrachromosomal elements are a reflection of the dynamic genomes of these organisms, and contain large numbers of transposable insertion elements (Table 1) (DasSarma et al., 2008). Several of the extrachromosomal elements also harbor large and small rRNA genes and have been designated as small (or mini-) chromosomes. The haloarchaeal genomes sequenced thus far, including extrachromosomal elements, range in size from 2.6 Mbp (*Halobacterium* sp. NRC-1) to 5.4 Mbp (*H. turkmenica*). The *Halobacterium* sp. NRC-1 genome consists of a 2 Mbp circular chromosome, and two large megaplasmids, pNRC200 (365 kbp) and pNRC100 (191 kbp). With the exception of *H. utahensis* and *H. walsbyi*, the other haloarchaeal genomes, all have

Table 1
Organization of haloarchaeal genomes.

Organism	Publication	Chromosomes (bp)	Extrachromosomal elements (bp)
Halobacterium sp. NRC-1	Ng et al. (2000)	2014239	pNRC200-365425 pNRC100-191346
Haloarcula marismortui	Baliga et al. (2004)	3131724	pNG700-410554 chrmII-288050 pNG600-155300 pNG500-132678 pNG400-50060 pNG300-39521 pNG200-33452 pNG100-33303
Natronomonas pharaonis	Falb et al. (2005)	2595221	pL131-130989 pL23-23486
Haloquadratum walsbyi	Bolhuis et al. (2006)	3132494	pL47-46867
Halobacterium sp. R-1	Pfeiffer et al. (2008)	2000962	pHS3-284332 pHS2-194963 pHS1-147625 pHS4-40894
Halorubrum lacusprofundi		2735295	chrmII-525943 pHLAC01-431338
Halogeometricum borinquense	Malfatti et al. (2009)	2820544	pHB500-362194 pHB400-339010 pHB300-210350 pHB200-194834 pHB100-17535
Halomicrobium mukohataei	Tindall et al. (2009)	3110487	pHM61-221862
Halorhabdus utahensis	Bakke et al. (2009)	3116795	
Haloferax volcanii	Hartman et al. (2010)	2847757	pHV4-635786 pHV3-437906 pHV1-85092 pHV2-6359
Haloterrigena turkmenica	Saunders et al. (2010)	3889038	pHT85-698495 pHT84-413648 pHT83-180781 pHT81-171943 pHT82-71062 pHT80-15815

similar genome arrangements with 2.0–3.9 Mbp chromosomes and large extrachromosomal elements (Table 1). *H. marismortui*, *H. turkmenica*, and *H. borinquense* have four megaplasmids, *Halobacterium* sp. R-1 has three, *H. volcanii* and *H. lacusprofundi* have two, and *N. pharaonis* and *H. mukohataei* have one, and several (*H. marismortui*, *H. volcanii*, *H. borinquense*, *N. pharaonis*, *H. turkmenica*, *H. walsbyi*, and *Halobacterium* sp. R-1) contain smaller plasmids ranging in size from 6 to 85 kbp. Only the *H. utahensis* genome is devoid of any plasmids, with all of its genetic material organized into a single 3.1 Mbp chromosome.

Here, we review the information transfer system of haloarchaea and highlight the similarities and differences among sequenced species representing ten genera, including biochemical and genetic studies primarily conducted on two model organisms, *Halobacterium* sp. NRC-1 and *H. volcanii*.

2. Origins of replication

DNA replication has been found to initiate at multiple distinct sites on the chromosome and large extrachromosomal elements in many archaeal organisms, but several have been shown to contain only a single origin (Berquist and DasSarma, 2008). Among the haloarchaea, replication origins have been experimentally mapped in *Halobacterium* sp. NRC-1 and *H. volcanii* (Norais et al., 2007; Berquist and DasSarma, 2008; Coker et al., 2009). In *Halobacterium* sp. NRC-1, whole genome marker frequency analysis (MFA) identified at least six peaks on the chromosome and extrachromosomal elements, consistent with the existence of multiple origins of replication. The two largest peaks on the chromosome, *oriC1* and *oriC2*, were upstream of eukaryotic-like *orc1/cdc6* replication genes, *orc7* and *orc10*, respectively. The *oriC1* origin was also identified genetically, via autonomous replicating activity and included *orc7*, upstream inverted repeats, and an intervening AT-rich sequence (Berquist and DasSarma, 2003). The genes around the *oriC1* region are syntenic among the sequenced haloarchaeal organisms (Fig. 1) (Coker et al., 2009). The *oriC1* inverted repeats were hypothesized to be origin binding elements, bound by Orc1/Cdc6 initiation factors that are homologs of the eukaryotic origin recognition complex subunit 1 and replicative helicase recruiter. A sequence similar to one copy of the inverted repeats was identified just upstream of the *orc1/cdc6* coding regions in each of the genomes, suggesting a role in *orc1/cdc6* gene regulation (Coker et al., 2009). Two origins, *oriP1* and *oriP2*, were identified on the extrachromosomal elements of *Halobacterium* sp. NRC-1. The *oriP1* origin is located between *orc9* and *repI* on the common region of pNRC100 and pNRC200. The *repI* gene is related to the *repH* gene family, which was shown to be required for replication of pNRC100 minireplicons commonly used as shuttle vectors in haloarchaea (Ng and DasSarma, 1993). An additional origin, *oriP2*, was identified near *orc4* on the larger extrachromosomal element, pNRC200.

The origins of replication have also been mapped in *H. volcanii* using a combination of replication ability assays and hybridization analysis (Norais et al., 2007).

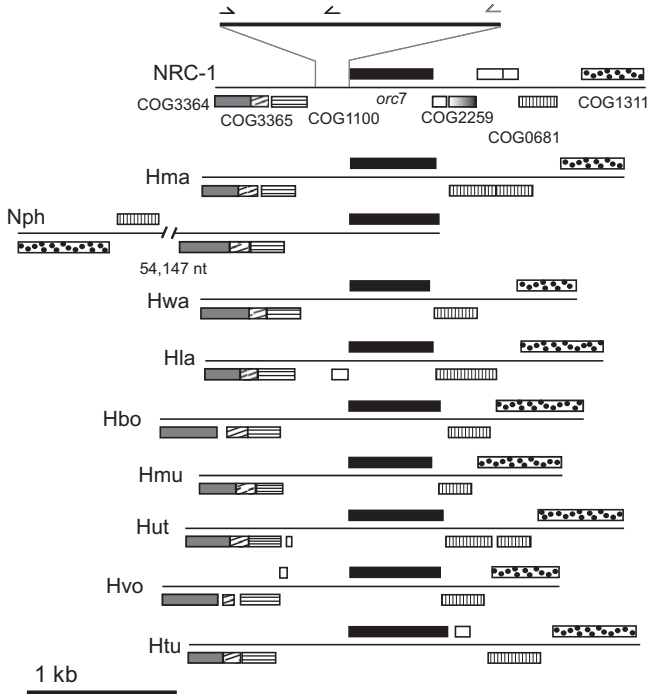

Fig. 1. Schematic representation of the highly conserved region of *Halobacterium* sp. NRC-1 *oriC1* origin of replication. Seven genes, the inverted repeats, as well as the putative regulatory sequence are conserved in all ten genomes. The genes conserved in all ten genomes are *orc7* (black) and genes associated with: COG3364, a zinc-ribbon containing protein (gray); COG3365, an uncharacterized archaeal protein (diagonal lines); COG1100, GTPase (horizontal lines); COG0681, signal peptidase I (vertical lines); and COG1311, family D DNA polymerase small subunit (polka-dots). Genes associated with COG2259, predicted membrane protein (gradient) and genes not associated with COGs (white) are not conserved.

Autonomously replicating sequences (ARS) were isolated from each of the large extrachromosomal elements and the chromosome, upstream of *orc10* on pHV1, *orc6* on pHV3, *orc3* on pHV4, and *orc1* and *orc5* on the chromosome. Hybridization analysis of the ARS associated near *orc10* on pHV1 suggested that this region is also found on pHV4 while the ARS associated near *orc3* was found to hybridize to the chromosome instead of pHV4. The discrepancies in origin location between sequence and hybridization data could be due to differences in the strains used in sequencing and genetic studies. Since *H. volcanii* was found to contain over 100 transposable insertion sequence (IS) elements, and the strain used in the genetic experiments was cured of pHV2 (the smallest *H. volcanii* plasmid), the possibility exists that other genomic rearrangements may also be present (Lam and Doolittle, 1989; Hartman et al., 2010).

In addition to the origin mapping studies in *Halobacterium* sp. NRC-1 and *H. volcanii*, the origins in several other archaea have been investigated. Similar to the haloarchaea, bioinformatics studies, two-dimensional gel analysis, and MFA of *Aeropyrum pernix* and two *Sulfolobus* spp., *S. acidocaldarius* and *S. solfataricus*, identified multiple sites of DNA replication initiation on each chromosome (Robinson et al., 2004; Lundgren and Bernander, 2007; Robinson and Bell, 2007; Duggin et al., 2008). By contrast, DNA replication has been shown to initiate at a single site on the *Pyrococcus abyssi* and *Methanothermobacter thermoautotrophicus* chromosomes (Myllykallio et al., 2000; Matsunaga et al., 2001).

3. DNA replication

Bioinformatic analysis showed that haloarchaeal DNA replication genes are highly conserved, and in some cases, a degree of complexity is seen compared to other archaea (Berquist and DasSarma, 2008). Most of the conserved genes are found on the haloarchaeal chromosomes, including those encoding nearly all DNA polymerases, replicative helicases, primases, and accessory proteins (Table 2). One of the most interesting findings was the expansion of the *orc1*/*cdc6* family, which correlated with the multiplicity of replication origins. Analysis of this gene family in sequenced haloarchaea showed the presence of 5–18 *orc*1/*cdc*6 homologs, while other archaeal genomes encode at most four (Fig. 2). The *orc1*/*cdc6* genes are found on all of the haloarchaeal chromosomes and many of the extrachromosomal elements, and likely play key roles in replication initiation of these replicons. However, eight of the haloarchaeal extrachromosomal elements lack any *orc1*/*cdc6* genes, similar to some other archaea, *e.g. Methanocaldococcus jannaschii* and *M. maripaludis* (Bult et al., 1996; Hendrickson et al., 2004). Three of these haloarchaeal replicons (*H. marismortui* pNG200 and pNG400 and *H. volcanii* pHV2) encode *rep* genes that have been shown to be involved in replication of mini-replicons of *Halobacterium* sp. NRC-1 (Ng and DasSarma, 1993). Interestingly, *H. borinquense* pHB100 and pHB300, *N. pharaonis* pL23, *H. walsbyi* pL47, and *H. turkmenica* pHT80 all lack both *orc1*/*cdc6* and *rep* genes; however, to date, their replication properties have not been studied.

The replicative minichromosome maintenance (MCM) helicase is coded by the *mcm* gene in eukaryotes and archaea. MCM helicases are hexameric enzymes that unwind DNA during chromosomal replication. They use energy from nucleoside triphosphate hydrolysis to translocate along one strand of the duplex DNA and displace the complementary strand (Patel and Picha, 2000). Structural studies using the thermophilic archaeon *A. pernix* proteins showed that Orc1/Cdc6 proteins bind to the inverted repeats at the origin of replication and the MCM helicase is positioned between the inverted repeats. The Orc1/Cdc6 domains are in opposite orientations, facing the AT-rich sequence (Kasiviswanathan et al., 2006; Gaudier et al., 2007). The directionality of Orc1/Cdc6 binding positions the MCM helicase protein near the AT-rich region and allows for DNA duplex melting and subsequent replisome formation. Each of the sequenced haloarchaea contains a copy of *mcm* on the chromosome, and *H. marismortui* and *H. turkmenica* contain a second copy on the pNG300 and pHT82 extrachromosomal elements, respectively. In eukaryotes, the active form of MCM is in complex with Cdc45 and the GINS complex. Although a homolog of Cdc45 has not been identified in any of the haloarchaea, homologs of GINS subunits have been found in each haloarchaeal genome examined.

Archaeal DNA primases are composed of two subunits homologous to the eukaryotic primase subunits, the small catalytic subunit (Pri1) and large regulatory subunit (Pri2), and are involved in the synthesis of oligoribonucleotide primers. *In vitro* studies of several archaeal primases have demonstrated both RNA and DNA synthesis without primers (Liu et al., 2001; Lao-Sirieix and Bell, 2004; Le Breton et al., 2007). Interestingly, all sequenced haloarchaeal genomes, like many other archaeal genomes, have, in addition to the eukaryotic primase genes, a gene homologous to the bacterial *dnaG* primase. While the precise role of DnaG in archaea has yet to be explored, it has been identified in exosome-like complexes, involved in RNA processing and degradation in *M. thermautotrophicus* and *S. solfataricus* (Farhoud et al., 2005; Walter et al., 2006). Genetic studies in *Halobacterium* sp. NRC-1 and *H. volcanii* have shown that the eukaryotic-like primase genes are essential (Berquist et al., 2007; Le Breton et al., 2007), while the *dnaG* gene appears to be dispensable (Le Breton et al., 2007).

Once the primers are synthesized, DNA synthesis can occur with recruitment of the sliding clamp loader, Rfc, homologous to the eukaryotic sliding clamp loader replication factor C (RFC) complex, DNA polymerase sliding clamp, Pcn, homologous to the eukaryotic PCNA (proliferating cell nuclear antigen), and DNA polymerase. The three-dimensional structure of the sliding clamp is very similar in all three domains of life, while diverged at the amino acid level. The bacterial β-clamp, eukaryotic PCNA, and archaeal Pcn all form pseudo-hexameric rings (six globular domains formed by two–three subunits) that accommodate double-stranded DNA through the pore (Winter et al., 2009). The sliding clamp tethers DNA polymerase to the DNA, increasing processivity. Interestingly, the subunit composition of archaeal Pcn differs and can form a hetero- or homotrimer. Three Pcn homologs have been identified in both *S. solfataricus* and *A. pernix*. While Pcn has been shown to be a heterotrimer in *S. solfataricus*,

Table 2
Conserved genes involved in DNA replication, chromatin structure and dynamics, DNA repair, and homologous recombination[a,b].

	COG	Number of homologs in NRC-1[c]	Number of homologs in Hma	Number of homologs in Nph	Number of homologs in Hwa	Number of homologs in Hla	Number of homologs in Hbo	Number of homologs in Hmu	Number of homologs in Hut	Number of homologs in Hvo	Number of homologs in Htu
DNA replication											
orc	COG1474: Cdc6-related protein, AAA superfamily ATPase	11 (10), 7	16, 9	5, 1	7	15, 10	9, 6	5, 1	5	14, 7	18, 8
rep	No associated COG	5 (3), 5	3, 3	0	1	2, 2	0	0	0	1, 1	1, 1
mcm	COG1241: Predicted ATPase involved in replication control, Cdc46/Mcm family	1	3, 1	2	1	2	1	2	1	1	2, 1
agh	COG1711: Uncharacterized protein conserved in archaea	1	1	1	1	1	1	1	1	1	1
pri1	COG1467: Eukaryotic-type DNA primase, catalytic (small) subunit	1	1	1	1	1	1	1	1	1	1
pri2	COG2219: Eukaryotic-type DNA primase, large subunit	1	1	1	1	1	1	1	1	1	1
dnaG	COG0358: DNA primase (bacterial type)	1	1	1	1	1	1	1	1	1	1
rfc	COG0470: ATPase involved in DNA replication	3	3	3	3	3	3	3	3	3	3
pcn	COG0592: DNA polymerase sliding clamp subunit (PCNA homolog)	1	1	1	1	1	1	1	1	1	1
polD1	COG1311: Archaeal DNA polymerase II, small subunit/DNA polymerase δ, subunit B	1	1	1	1	1	1	1	1	1	1
polD2	COG1933: Archaeal DNA polymerase II, large subunit	1	1	1	1	1	1	1	1	1	1
polB	COG0417: DNA polymerase elongation subunit (family B)	2, 1	2, 1	1	1	1	1	1	2	2, 1	1
rnhA	COG0328: Ribonuclease HI	1	2, 1	1	3	1	3, 2	2	1	2, 1	1
rnhB	COG0164: Ribonuclease HII	1	1	1	1	1	1	1	1	1	1
ligI	COG1793: ATP-dependent DNA ligase	1	1	1	2	1	1	1	0	1	1
ligA	COG0272: NAD-dependent DNA ligase (contains BRCT domain type II)	0	1	1	1	1	1	1	1	1	1
lig	COG1423: ATP-dependent DNA ligase, homolog of eukaryotic ligase III	0	1	0	0	0	1, 1	0	1	0	1
rfa1, *2*, *3*, & *6*	COG1599: Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit and related ssDNA-binding proteins	5, 2	5, 2	3	4	3	3	3	4	6, 3	4, 2
rfa7 & *8*	COG3390: Uncharacterized protein conserved in archaea	2	3, 1	2	2	2	2	2	2	2	2, 1
ral	No associated COG	1	1	1	1	1	1	1	1	1	1
gyrA	COG0188: Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit	1	1	1	1	1	1	1	1	1	1
gyrB	COG0187: Type IIA topoisomerase (DNA gyrase/ topo II, topoisomerase IV), B subunit	1	1	1	1	1	1	1	1	1	1
topA	COG0550: Topoisomerase IA	1	1	1	1	1	1	1	1	1	1
top6A	COG1697: DNA topoisomerase VI, subunit A	1	1	1	1	1	1	1	1	1	1
top6B	COG1389: DNA topoisomerase VI, subunit B	1	1	1	1	1	1	1	1	1	1
Chromatin structure and dynamics											
hpyA	COG2036: Histones H3 and H4	1	1	1	1	1	1	1	1	1	1
nhp	No associated COG	1	1	1	1	1	1	1	1	1	1
act2	No associated COG	1	1	1	1	1	1	1	1	1	1
act5	COG1243: Histone acetyltransferase	1	1	1	1	1	1	1	1	1	1
act1, *3*, 4, *6-10* & *pai1*	COG0454: Histone acetyltransferase HPA2 and related acetyltransferases	8	11, 1	6	5	8	9	10	5	9	15, 3

(*continued on next page*)

Table 2 (*continued*)

	COG	Number of homologs in NRC-1[c]	Number of homologs in Hma	Number of homologs in Nph	Number of homologs in Hwa	Number of homologs in Hla	Number of homologs in Hbo	Number of homologs in Hmu	Number of homologs in Hut	Number of homologs in Hvo	Number of homologs in Htu
hdf	COG0123: Deacetylases, including yeast histone deacetylase and acetoin utilization protein	1	1	1	1	1	1	1	1	1	1
DNA repair and homologous recombination											
phr	COG0415: Deoxyribodipyrimidine photolyase	2	2	2	2	2	1	2	2	2	1, 1
uvrA	COG0178: Excinuclease ATPase subunit	1	1	1	1	1	1	1	1	1	1
uvrB	COG0556: Helicase subunit of the DNA excision repair complex	1	1	1	1	1	1	1	1	1	1
uvrC	COG0322: Nuclease subunit of the excinuclease complex	1	1	1	1	1	1	1	1	1	1
uvrD	COG0210: Superfamily I DNA and RNA helicases	1	1	1	1	1	2, 1	1	1	1	1
rad2	COG0258: 5'-3' exonuclease (including N-terminal domain of PolI)	1	1	1	1	1	1	1	1	1	1
rad3	COG1199: Rad3-related DNA helicases	2	2	2	2	2	2	2	2	2	2
rad25a & *b*	COG1061: DNA or RNA helicases of superfamily II	3 (2), 2	2	3	3	2	2	2	1	4, 1	2, 1
bax1	COG3372: Uncharacterized conserved protein	2 (1), 2	1	2	2	3, 2	1	1	1	3, 1	1
mutS1	COG0249: Mismatch repair ATPase (MutS family)	2	2	2	2	2	2	2	2	2	2
mutS2	COG1193: Mismatch repair ATPase (MutS family)	1	1	1	1	1	1	1	1	1	1
mutL	COG0323: DNA mismatch repair enzyme (predicted ATPase)	1	1	1	1	1	1	1	1	1	1
ogg & *alkA*	COG0122: 3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase	2	2	2	1	2	2	2	2	2	2
mutY	COG1194: A/G-specific DNA glycosylase	1	1	1	1	1	2	1	1	1	1
nthA	COG0177: Predicted EndoIII-related endonuclease	2	3, 1	2	2	2	2	3	2	2	2
nthB	COG1833: Uncharacterized conserved protein	1	1	1	1	0	1	1	1	1	1
nfi	COG1515: Deoxyinosine 3'endonuclease (endonuclease V)	1	1	1	1	1	1	1	1	1	1
xthA	COG0648: Endonuclease IV	1	1	1	1	1	1	1	1	1	1
urg	COG1573: Uracil-DNA glycosylase	3	3	2	2	3	3	3	3	3	3
mtl, *apa*, & *mutT*	COG0494: NTP pyrophosphohydrolases including oxidative damage repair enzymes	6	6	5	5	5	7, 1	7	6	6	4
ogt	COG0350: Methylated DNA–protein cysteine methyltransferase	1	1	1	1	1	1	1	1	1	1
sod	COG0605: Superoxide dismutase	2	1	1	1	2	1	2, 1	2	2, 1	1
mre11	COG0420: DNA repair exonuclease	1	1	1	1	1	1	1	1	1	1
rad50	COG0419: ATPase involved in DNA repair	1	1	1	1	1	1	1	1	1	1
umuC	COG0389: Nucleotidyltransferase/DNA polymerase involved in DNA repair	1	1	1	1	1	1	1	1	1	1
radA	COG0468: RecA/RadA recombinase	2	2	2	2	2	2	2	2	2	2
recJ	COG0608: Single-stranded DNA-specific exonuclease	1	1	1	1	1	1	1	1	1	1
arj	COG1107: Archaea-specific RecJ-like exonuclease, contains DnaJ-type Zn finger domain	2	2	2	2	2	2	2	2	2	2
hjr	COG1591: Holliday junction resolvase - archaeal type	1	1	1	1	1	1	1	1	1	1
rbl1 & *3*	COG1637: Predicted nuclease of the RecB family	2	1	1	1	1	1	1	1	1	1

[a] When two numbers are expressed, the first number is the total number of genes, and the second number is the number of genes coded on extrachromosomal elements.

[b] Organism name abbreviations are as follows: *Halobacterium* sp. NRC-1 (NRC-1), *H. marismortui* (Hma), *N. pharaonis* (Nph), *H. walsbyi* (Hwa), *H. lacusprofundi* (Hla), *H. borinquense* (Hbo), *H. mukohataei* (Hmu), *H. utahensis* (Hut), *H. volcanii* (Hvo), and *H. turkmenica* (Htu).

[c] When three numbers are expressed, the first number is the total number of genes, the number in parentheses is the number of unique genes, and the third number is the number of genes coded on extrachromosomal elements.

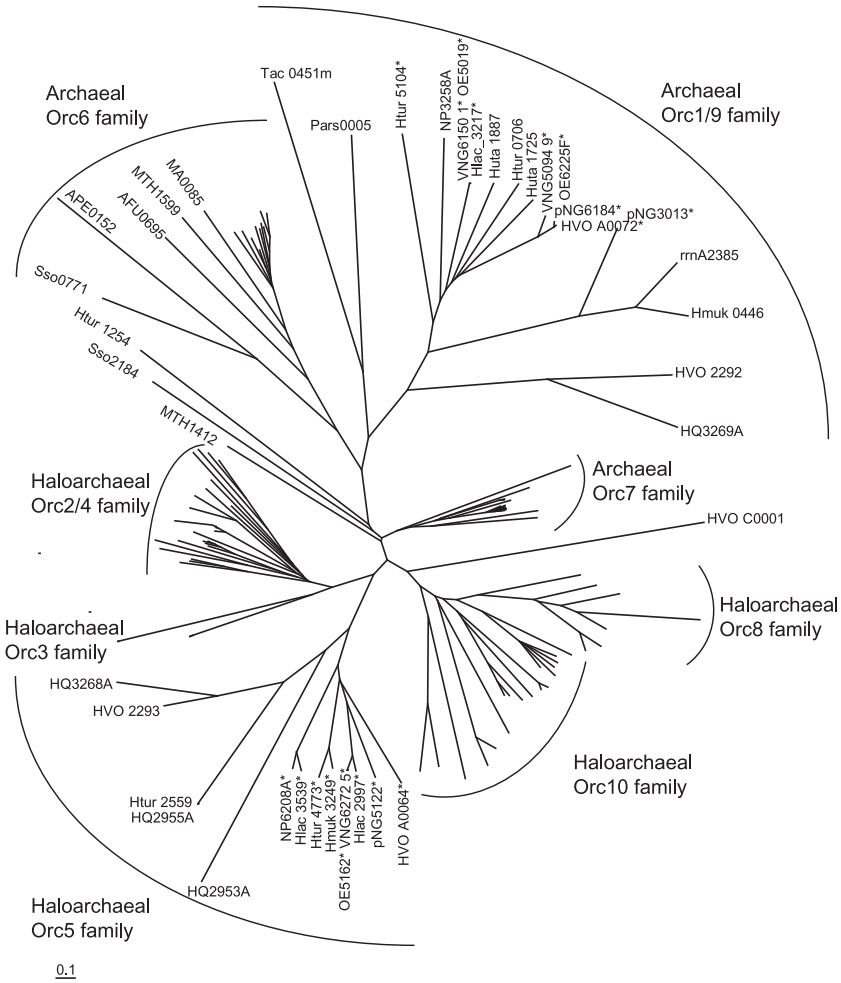

Fig. 2. Orc Protein Clusters. Neighbor-joining radial tree of truncated haloarchaeal sequences and other archaeal sequences in COG1474 (*Archaeoglobus fulgidus*, *Aeropyrum pernix*, *Methanosarcina acetivorans*, *Pyrococcus abyssi*, *Pyrobaculum arsenaticum*, *Pyrococcus horikoshii*, *Sulfolobus solfataricus*, and *Thermoplasma acidophilum*). *indicates proteins coded on extrachromosomal elements. Haloarchaeal Orc2/4 family includes VNG6164*, VNG6363*, rrnAC1053, pNG1024*, HVO_1725, HVO_C0057*, HVO_B0001*, HVO_A0257*, Hbor_14700, Hbor_37410*, Hbor_34470*, Hbor_33490*, Huta_1077, Hlac_1524, Hlac_3512*, Hlac_3641*, Hlac_2833*, Htur_0510, Htur_1155, Htur_1843, Htur_2434, Htur_3152, Htur_5222*, Htur_5244*, Htur_3916*, and OE6037*. Haloarchaeal Orc3 family includes VNG6187*, pNG6174*, HQ2959A, Htur_4912*, Htur_3967*, and OE5069*. In addition to those shown, archaeal Orc6 family also includes VNG2271, rrnAC2862, HVO_0194, Hbor_00440, Hmuk_1003, Huta_1055, Hlac_2631, NP0588A, HQ3680A, Htur_3634, and OE4184. Archaeal Orc7 family includes VNG2411, rrnAC2711, HVO_0001, Hbor_02110, Hmuk_0815, Huta_1613, Hlac_0001, NP0596A, HQ1001A, Htur_0001, OE4380, Sso0257, APE0475, Mac0001, Pho 0124, PAB2265, AFU0244, and Ta0636. Haloarchaeal Orc8 family includes VNG1224, rrnAC1569, Hbor_31040*, Hmuk_1814, Hlac_1085, Hlac_1078, NP3096A, and OE2753. Haloarchaeal Orc10 family includes VNG0045, rrnAC1262, rrnAC1568, pNG5027*, rrnB0004*, rrnB0063*, pNG7187*, HVO_0634, HVO_2042, HVO_A0001*, Hbor_36740*, Hbor_37550*, Hlac_3320*, Hlac_3367*, Hlac_2747*, Hlac_2958*, Htur_5210*, OE1076F, and OE6288R*.

the *A. pernix* subunits can form both hetero- and homotrimers (Daimon et al., 2002; Williams et al., 2006). In most other archaea, including all of the sequenced haloarchaea, one Pcn homolog is present and its homotrimeric structure has been recently confirmed by the crystal structure of the *H. volcanii* Pcn (Winter et al., 2009).

The ring-structure of the sliding clamp must open and close to load onto DNA, which is accomplished by a clamp loader. Similar to the sliding clamp, the overall structure of the clamp loader is conserved in all three domains; however, the subunit composition is remarkably different. The complex is composed of three different subunits in bacteria, five in eukaryotes, and two in most archaea (Lao-Sirieix et al., 2007). Studies of the archaeal clamp loader complex have shown that the Rfc complex enhances the activity of DNA polymerase. However the composition of the complex differs, with a eukaryotic-like pentameric complex of one large and four small subunits in *Archaeoglobus flugidus* and *S. solfataricus* (Pisani et al., 2000; Seybert et al., 2002), a trimer or hexamer complex in *P. abyssi* (Henneke et al., 2002), and a hexamer in *M. thermoautotrophicus* (Kelman and Hurwitz, 2000). Interestingly, molecular analysis of the *Methanosarcina acetivorans* clamp loader resulted in identification of a three subunit complex with a 3:1:1 ratio (RfcS1:RfcS2:RfcL). This unique subunit composition may represent an intermediate form between other archaeal

and eukaryotic complexes (Chen et al., 2005). As in *M. acetivorans*, genes coding for two small and one large Rfc subunits were identified in each haloarchaeal genome, suggesting that the haloarchaeal Rfc complex composition may have a similar composition and ratio.

Each of the haloarchaea encode two families of DNA polymerases, the single subunit family B polymerase, PolB, and the euryarchaeal two-subunit family D polymerase consisting of PolD1 (small subunit) and PolD2 (large subunit). While most have a single copy of each enzyme family, four haloarchaea (*Halobacterium* sp. NRC-1, *H. marismortui*, *H. volcanii*, and *H. utahensis*) have two copies of the PolB family genes, with one copy on extrachromosomal elements in *Halobacterium* sp. NRC-1, *H. marismortui*, and *H. volcanii* (pNRC200, pNG600, and pHV4, respectively). Genetic studies conducted in *Halobacterium* sp. NRC-1 have found that genes encoding both PolD subunits and only the chromosomal PolB gene are essential (Berquist et al., 2007). In *P. abyssi*, biochemical work showed that PolD is able to utilize RNA primers, while PolB requires DNA primers. This has given rise to the hypothesis that PolB serves as the leading strand and the PolD as the lagging strand polymerase (Henneke et al., 2005).

After replication, the RNA primers must be removed and gaps sealed. In all three domains, RNase H enzymes cleave the RNA strand in RNA/DNA hybrids. The RNA is then cleaved from the newly synthesized DNA at the RNA–DNA junction by DNA polymerase I in bacteria and the flap endonuclease in eukaryotes and archaea. Archaea usually encode a type II RNase H, similar to both of the bacterial RNase HII and RNase HIII enzymes and eukaryotic RNase H2, and lack a type I homolog. However, each haloarchaeal genome possesses at least a single copy of the type I (RnhA) and type II (RnhB) RNase H genes. The second type I RNase H gene in the *H. marismortui* and *H. volcanii* genomes and two of the three type I RNase H genes in the *H. borinquense* genome are located on extrachromosomal elements pNG700, pHV4, and pHB400, respectively. In addition to cleaving a RNA/DNA hybrid at multiple sites, the *Halobacterium* sp. NRC-1 RNase HI enzyme was found to cleave at the RNA–DNA junction, unlike the RNase HI from *E. coli* (Ohtani et al., 2004). This finding is interesting as it suggests that the haloarchaeal RNase HI may also cleave at the RNA–DNA junction *in vivo* like the flap endonuclease.

Okazaki fragments on the lagging-strand are joined in archaea by the eukaryotic-like ATP-dependent DNA ligase, LigI. A bacterial-type NAD^+-dependent ligase was also identified, first in *H. marismortui* (Berquist et al., 2005), and subsequently in all of the haloarchaea, except *Halobacterium* sp. NRC-1. An additional ATP-dependent DNA ligase, homologous to the eukaryotic ligase III, is present on the chromosomes of three of the haloarchaea (*H. marismortui*, *H. utahensis* and *H. turkmenica*), and on pHB500 in *H. borinquense*. Additional genetic studies have shown that the ATP- and NAD^+-dependent ligases are not essential under standard laboratory conditions in *H. volcanii*, as determined by single-gene deletions. An attempt to form a double-deletion of these genes was unsuccessful, indicating that their products share an essential ligase function, with each gene able to compensate for the loss of the other (Zhao et al., 2006).

During DNA replication, single-stranded DNA is bound by single-stranded DNA binding proteins. In eukaryotes these proteins, termed RPA70, RPA32, and RPA14, form a complex (replication factor/protein A complex) (Zou et al., 2006). Three distinct homologs to the eukaryotic RPA70 and two distinct homologs to RPA32 are conserved in all of the sequenced haloarchaeal genomes. A homolog of the eukaryotic RPA14 has not, however, been identified in any of the sequenced haloarchaeal genomes. In addition to DNA replication, RPA proteins and their homologs in archaea have also been found to play critical roles in DNA repair, and are required for survival after high-energy ionizing radiation damage (DeVeaux et al., 2007).

DNA topoisomerases function to regulate the DNA superhelical density and decatenate interlinked DNA molecules. Haloarchaea encode three separate topoisomerases: the two subunits of the bacterial-like DNA gyrase, the type II topoisomerase VI, and a type I topoisomerase (IA). DNA gyrases are type II enzymes that introduce negative supercoiling into relaxed or supercoiled DNA, and by analogy to *E. coli*, archaeal DNA gyrases are thought to reverse positive supercoiling that occurs ahead of the replication fork, and may be responsible for the global DNA topology in archaeal genomes (Forterre and Gadelle, 2009). An additional factor supporting the idea that DNA gyrase may be responsible for regulating the global DNA topology in archaeal genomes is that microorganisms that carry gyrase genes have negatively supercoiled DNA and archaea that lack gyrase genes have more relaxed DNA (Forterre and Gadelle, 2009). Structural studies of archaeal topoisomerase VI subunits in *M. jannaschii* and *Sulfolobus shibatae* have led to a proposed mechanism for topoisomerase VI in positive supercoiling removal and decatenation similar to DNA gyrase, suggesting a possible role in chromosome segregation (Nichols et al., 1999; Corbett and Berger, 2003). Studies of the *S. acidocaldarius* cell cycle and cell division machinery did not support this mechanism however, indicating that topoisomerase VI may perform a different role in archaeal cells (Lundgren and Bernander, 2007; Lindas et al., 2008). Transcriptional mapping of *S. acidocaldarius* cell cycle showed constitutive expression of the topoisomerase VI genes throughout the cell cycle (Lundgren and Bernander, 2007), and the cell division machinery proteins were expressed, indicating chromosome segregation was completed when topoisomerase VI was inhibited (Lindas et al., 2008). The conflicting results indicate the role of topoisiomerase VI has yet to be fully understood. Interestingly, the DNA gyrase and topoisomerase VI genes are located adjacent to each other and in opposite orientation in six of the ten haloarchaeal genomes (*Halobacterium* sp. NRC-1, *H. volcanii*, *H. walsbyi*, *H. marismortui*, *H. utahensis* and *H. mukohataei*) which suggests that these topoisomerases may be coordinately regulated. While the function of archaeal IA topoisomerases is still not clearly established, phylogenetical analysis showed that they are similar to other type IA topoisomerases and likely function to partially relax negative, but not positive, supercoiling (Forterre et al., 2007).

4. Chromatin structure and dynamics

Unlike those in DNA replication, the genes coding for archaeal proteins involved in chromatin structure are quite diverse and have a more limited phylogenetic distribution. In haloarchaea, nucleosome-like structures were originally identified by examining electron microscopic (EM) images of *Halobacterium* chromosome fibers (Shioda et al., 1989). The EM images showed two distinct species: fine fibers that were later shown to be protein-free DNA, and nucleosome-like particles later identified as protein-associated DNA (Takayanagi et al., 1992). The archaeal histone protein, Hpy, forms the nucleosome-like structures observed in the electron microscope, which are similar to the eukaryotic H3/H4 tetramer. *hpy* genes are found in most euryarchaea as well as in *Cenarchaeum symbiosum*, and have been identified in all sequenced haloarchaeal genomes (Table 2). Unlike most other euryarchaea, which have up to seven *hpy* genes, haloarchaea only have a single *hpy* gene. Hpy proteins form homo- or heterodimers and combine into tetramers with ~90 bp of DNA wrapping around the complex (Samson and Reeve, 2007). Originally identified in *Methanosarcina* and named the methanogen chromatin protein 1 (MC1), a gene coding a non-histone chromosomal protein, Nhp, has also been identified in all sequenced haloarchaeal genomes. MC dimers have been shown to kink DNA by up to 120° (Samson and Reeve, 2007).

Several additional archaeal chromatin binding proteins, including Alba, Sul7, 7kMk, and the HU-like proteins have been studied, but none of them are present in any of the haloarchaeal genomes. Alba (acetylation lowers binding affinity) proteins dimerize, and each monomer interacts with DNA, bridging two regions of the chromosome. Genes coding Alba are found in most non-haloarchaeal genomes. Also absent is Sul7, which is found only in *Sulfolobus* species, and which binds and bends DNA up to 60° (Luijsterburg et al., 2008), and 7kMk, which is unique to *Methanopyrus*. Homodimers of 7kMk bind DNA non-specifically and form loop-like structures, but the precise mechanism of DNA compaction has yet to be determined (Samson and Reeve, 2007). Some archaea also contain HU-like proteins, identified in *Thermoplasma*, and are homologous to and thought to function like bacterial HU proteins, which bind and bend DNA up to 160° (White and Bell, 2002; Luijsterburg et al., 2008).

Even though archaeal histone proteins lack tails used for covalent modification of eukaryotic histones, post-translational modification of chromatin proteins do play a role in regulating archaeal DNA compaction. For example lysine modification has been demonstrated for the MC1 protein in *Methanosarcina mazei* (Manzur and Zhou, 2005) and has also been observed with *Saccharomyces cerevisiae* core histone protein H4 (Hyland et al., 2005). No less than seven acetyltransferase (*act*) genes and a deacetylase (*hdf*) gene are conserved in all ten sequenced haloarchaeal genomes and may be involved in post-translational modification of HpyA and Nhp (Table 2). Several Act protein clusters (Act1, 3, 4, and 6–10) as well as Pai1 are members of the histone acetyltransferase protein family (COG0454). One *H. marismortui* and three *H. turkmenica* homologs in this expanded gene family are located on extrachromosomal elements (chromosome II and pHT85, respectively). Two of the *H. volcanii* Act family homologs (*act9* cluster, HVO_1821 and *act10* cluster, HVO_1756, COG0454) as well as the *act5* homolog (HVO_2888, associated with COG1243) have been studied genetically, and found to be non-essential (Altman-Price and Mevarech, 2009). Interestingly, the *act9* and *act10* family genes could be deleted simultaneously as could the *act5* and *act10* family genes. However, the *act5* and *act9* double-deletion was lethal. An additional conserved acetyltransferase gene, *act2*, encodes an ArgA-like acetyltransferase. Two of the *H. turkmenica* orthologs, one *act1* and one *act10* family gene, are located on pHT85. The conserved deacetylase family (HVO_0522) gene in *H. volcanii* was also targeted for deletion, and was found to be essential (Altman-Price and Mevarech, 2009).

5. DNA repair and homologous recombination

DNA repair is crucial to the viability of haloarchaeal cells because damage due to environmental factors such as UV light and radiation is a major challenge in their environments. Moreover, desiccation has also been shown to result in double-stranded DNA breaks and haloarchaeal survival in hypersaline and dry environments may have contributed to their ability to repair such DNA lesions (DasSarma et al., 2001). A number of studies on DNA repair genes has been conducted, primarily on the *Halobacterium* sp. NRC-1 and *H. volcanii* models, and provided insights into the underlying processes (Berquist et al., 2007). The importance of these genes is underscored by their high degree of conservation in haloarchaea, with over 90% of the DNA repair and homologous recombination genes conserved in nearly all sequenced genomes (Table 2).

In haloarchaea, a key mechanism for repair of UV damage prevalent in their natural environment is direct photorepair (Hescox and Carlberg, 1972; McCready and Marcello, 2003). Genetic and biochemical analysis of *Halobacterium* sp. NRC-1 showed that one of the two deoxyribodipyrimidine photolyases, Phr2, is responsible for the primary cyclobutane pyrimidine dimer (CPD) photolyase repair activity, but not the activity necessary for repair of 6–4 photoproducts (McCready and Marcello, 2003). By contrast, the function of Phr1 is not known, though it has been proposed to function as a cryptochrome, a blue-light receptor protein involved in circadian rhythms (Kanai et al., 1997). Though, the *phr2* photolyase is conserved in all haloarchaeal genomes, the *phr1* is not completely conserved. Interestingly, the *H. turkmenica*, the *phr2* gene is coded on the extrachromosomal element pHT82.

Dark repair of UV lesions in bacteria occurs by nucleotide excision repair and requires the UvrA protein for recognition, UvrB and UvrC for excinucleolytic incisions on either side of the lesion, and UvrD, a DNA helicase II, for removal of the damaged strand. UvrABCD initiates repair of CPDs and 6–4 photoproducts as well as other bulky lesions. All of the haloarchaea possess a single copy of these four

bacterial-like repair genes, except *H. borinquense*, which has two *uvrD* genes, one on the chromosome and another on pHB100. Most non-halophilic archaea lack the UvrABCD nuclease altogether. UvrA, C, and AC double mutants showed a decreased survival rate and reduced CPD repair after UV irradiation in *Halobacterium* sp. NRC-1 (Crowley et al., 2006). Surprisingly, the gene for *uvrD* could be deleted in this strain with no observed change in tolerance to the strong alkylating agent MNNG (Busch and DiRuggiero, 2010). In addition to the bacterial-like repair genes, all ten haloarchaea have eukaryotic-like repair genes, including *rad2, rad3,* and *rad25*. One of the four copies of *rad25* in *H. volcanii* and one of the two copies in *H. turkmenica* are located on the extrachromosomal elements pHV4 and pHT81, respectively, and two of the three copies in *Halobacterium* sp. NRC-1 are located on the common region of pNRC100 and pNRC200. For Rad2, a member of the Rad2/FEN1 family of flap endonucleases, work in *S. cerevisiae* indicates that it makes an incision on the 3′ side of the UV photoproduct.

Archaeal Rad3 and Rad25 correspond to human XPD and XPB, respectively. They are involved in both NER and transcription initiation as part of the DNA helicase subunits of TFIIH in higher organisms. Recent investigations of the eukaryotic factors showed that the XPD helicase activity is required for DNA opening and the XPB helicase activity is not (Coin et al., 2007). The XPD ATPase activity is necessary for DNA repair, while the ATPase activity of XPB is involved in recruitment of the other subunits of TFIIH, including XPD. Since archaea only possess homologs to the two helicase subunits, it was unclear if they would perform functions analogous to their eukaryotic homologs. Recent studies in *S. acidocaldarius* and *S. solfataricus* have shown that the archaeal XPD and XPB homologs are indeed involved in NER with roles similar to their eukaryotic counterparts (Rouillon and White, 2010; Rudolf et al., 2010). The XPD homolog Rad3 binds and unwinds DNA (Rudolf et al., 2010). The XPB homolog Rad25, in complex with a novel nuclease Bax1, unwinds and cleaves at small DNA bubbles. The complex acts on the 5′ side of the DNA bubble, and given the 3′–5′ polarity of XPB, it is likely that the archaeal XPB and XPD homologs work in the same direction to open the bubble (Rouillon and White, 2010). Bax1 homologs are present in all of the sequenced haloarchaea and are coded adjacent to *rad25* in *Halobacterium* sp. NRC-1, *H. volcanii*, *N. pharaonis*, *H. walsbyi*, and *H. turkmenica*.

Most mismatches that result from DNA polymerase errors during replication are the target of the post-replicative DNA mismatch repair (MMR) system. The strand-specific mismatch repair system has been best characterized in *E. coli* and many of the genes involved are conserved among the haloarchaeal genomes. Three proteins, MutS, MutL, and MutH, carry out two essential functions: mismatch detection and targeting repair to a specific strand. In *E. coli*, MutS recognizes an error of 1–4 nucleotides on one strand, and has weak ATPase activity, which may play a role in mismatch recognition, as well as signaling other MMR proteins to assemble in the repair complex (Heinze et al., 2009). MutL then binds to the MutS–DNA complex and acts as a molecular switch in which nucleotide binding modulates interactions with and enhancement of other MMR proteins (Kunkel and Erie, 2005). One *mutL* gene and three distinct *mutS* genes are found in each haloarchaeal genome, two (*mutS1a* and *mutS1b*) are associated with one MutS protein family (COG0249) and the third (*mutS2*) is associated with another (COG1193). Genetic studies in *Halobacterium* sp. NRC-1 targeted *mutS1a*, *mutS1b*, and *mutL*, and all three genes were reported to be non-essential (Busch and DiRuggiero, 2010). In *E. coli*, MutH, activated by MutL, cleaves the newly synthesized unmethylated strand at hemimethylated GATC sites near the mismatch (Kunkel and Erie, 2005). Genes coding for MutH are not found in any of the haloarchaeal genomes, suggesting that a novel unidentified protein must carry out the endonuclease activity of MutH in haloarchaea.

Oxidative DNA damage resulting from reactive oxygen species (ROS) is mainly repaired by base excision repair (BER) in all organisms studied, and most likely also in haloarchaea. During BER, a DNA glycosylase cleaves the *N*-glycosidic bond of the damaged base and nicks the DNA strand in a lyase reaction, followed by nucleotide replacement and religation. The helix–hairpin-helix (HhH) superfamily of BER glycosylases is comprised of a group of enzymes that specifically recognize and excise damaged bases, and their amino acid sequences are thought to convey specificity in damage recognition (Denver et al., 2003). The HhH superfamily includes A/G-specific adenine glycosylases (MutY), alkyladenine glycosylase (AlkA), and endonuclease III (Nth). The *mutY* and *nth* genes are conserved throughout all the sequenced haloarchaeal genomes, while *alkA* is missing in *H. walsbyi*. There is an expansion of the *nthA* genes, with three genes identified in *H. marismortui* and *H. mukohataei* and two in the remaining genomes. One of the three *nthA* genes in *H. marismortui* is found on the large pNG700 extrachromosomal element. While *nthA* genes are found in all sequenced haloarchaeal genomes, a related gene, *nthB*, is missing in the *H. lacusprofundi* genome.

Conserved across all haloarchaeal genomes, the oxoguanine glycosylase (*ogg*) gene encodes the enzyme that removes 7,8-dihydro-8-oxoguanine (8-oxoG), which causes GC to TA transversions in replication (Robey-Bond et al., 2008). AlkA and Ogg are both in the 3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase COG (COG0122). The *nfi* gene codes for endonuclease V, a magnesium-dependent enzyme that makes a nick at the second phosphodiester bond 3′ to the substrate lesion in *E. coli*. In archaea, endonuclease V also has the ability to act on DNA containing deaminated bases as well as other DNA damage sites formed by intracellular ROS that can alter the secondary structure of DNA (Kanugula et al., 2005). The *xthA* gene codes for endonuclease IV, which is an enzyme that recognizes AP sites of double-stranded DNA and cleaves the phosphodiester bond 5′ to the damaged portion, thus generating a hydroxyl group at the 3′-terminus (Kiyonari et al., 2009a; Kiyonari et al., 2009b). The genes that code for uracil-DNA glycosylases (*urg*) catalyze the removal of uracil by moving it from the double helix to their binding pockets where the glycosidic bond is hydrolyzed by a water molecule activated by a polar amino acid (Sartori et al., 2002). The *nfi, xthA,* and *urg* genes are all

conserved among the haloarchaeal genomes, and there is an expansion of the *urg* genes, with two encoded in the *N. pharaonis* and *H. walsbyi* genomes and three in the remaining eight genomes.

In *E. coli* the MutT proteins hydrolyze 8-oxo-dGMP from the nucleotide pool (Maki and Sekiguchi, 1992), and haloarchaeal MutT and MutT-like proteins (Mtl) are likely involved in the elimination of oxidized bases in the dNTP pool, although their substrates are unknown. While only *Halobacterium* sp. NRC-1 and *H. borinquense* encode *mutT* genes, three to five *mtl* genes are encoded by each organism. Furthermore, MutT, the Mtl homologs, and Apa (diadenosine tetraphosphate pyrophosphohydrolase) are all associated with a single protein cluster (COG0494) and together form an expanded family in sequenced haloarchaea with each organism encoding up to seven of these genes (Table 2). The *mutT* gene in the *H. borinquense* genome is encoded on the extrachromosomal element pHB400. All ten haloarchaea contain an *ogt* gene, coding a 6-*O*-methylguanine-DNA methyltransferase homolog, likely involved in DNA repair and protection against DNA alkylation damage. The superoxide dismutase gene (*sod*) encodes the enzyme that is responsible for limiting the cytotoxic damage from ROS, which is induced by extreme conditions such as oxidative stress and excessive irradiation (May and Dennis, 1989; Cheeseman et al., 1997). The haloarchaeal Sod proteins are all closely related, with multiple copies of the gene present in *Halobacterium* sp. NRC-1, *H. volcanii*, *H. mukohataei*, *H. utahensis*, and *H. lacusprofundi*, and a single copy found in the remaining five genomes. The second *sod* gene copies in *H. volcanii* and *H. mukohataei* are located on the extrachromosomal elements, pHV4 and pHM61.

Maintaining genomic integrity through the repair of DNA double-stranded breaks (DSBs) is critical to the survival of the haloarchaeal cells, especially following ionizing radiation which produces extensive DNA fragmentation (Daly et al., 1994; DiRuggiero et al., 1997). In eukaryotes, the Mre11–Rad50 complex, termed the MR complex, performs the role of DSB recognition. However, in *Halobacterium* sp. NRC-1, Mre11 is hypothesized to act as a sensor for DNA DSBs as well as a nuclease for the generation of single-stranded templates for recombinase activity. The ATP-dependent DNA binding activity of Rad50 (Hopfner et al., 2000) is reportedly not required for recombinational repair of DNA DSBs in *Halobacterium* sp. NRC-1 (Kish and DiRuggiero, 2008). Both the *mre11* and *rad50* genes are conserved among the haloarchaeal genomes. In *Halobacterium* sp. NRC-1, these genes were reported to be non-essential, as each gene could be knocked out individually and together. Deleting *mre11* resulted in a decrease in DSB repair that was not observed in the *rad50* deletion. These findings suggest a difference in the roles of the MR complex in archaea as compared to the eukaryotic model (Kish and DiRuggiero, 2008).

Single-stranded DNA (ssDNA) binding proteins are involved in DNA replication, repair, and recombination. Among eukaryotes, the RPA complex has been shown to be essential for MMR, BER, and DSB repair (Zou et al., 2006). It is probable that in haloarchaea, which contain genes homologous to the eukaryotic genes involved in these repair systems, the RPA homologs are also essential. Among the sequenced haloarchaea three distinct RPA70 homologs (Rfa1, 2, and 3) are present. One of the two *rfa3* homologs in *H. marismortui* and the *rfa2* homolog in *H. turkmenica* are coded on the extrachromosomal elements, pNG200 and pHT81, respectively. A gene coding for a fourth RPA70 homolog (Rfa6) is present on the chromosome of *H. utahensis* and *H. walsbyi* and on the extrachromosomal elements pNG600 and pHT83 in *H. marismortui* and *H. turkmenica*, respectively. Two *rfa6* genes in *Halobacterium* sp. NRC-1 are encoded on pNRC100 and pNRC200, and three *rfa6* genes in *H. volcanii* are encoded on pHV4. Genes coding for two distinct RPA32 homologs (Rfa7 and 8) are present in all of the sequenced haloarchaeal genomes. The *rfa7* gene in *H. turkmenica* and one of the *rfa8* genes in *H. marismortui* are located on the extrachromosomal elements, pHT81 and pNG200. In *Halobacterium* sp. NRC-1, *rfa3* and *rfa8* are transcriptionally linked, forming an operon with a third gene, *ral*, a conserved haloarchaeal protein which has been hypothesized to be the third subunit of the RPA complex (DeVeaux et al., 2007). However, Ral lacks sequence similarity to RPA14 and does not contain any OB fold domains usually found in RPA proteins. Transcriptome analysis of highly ionizing radiation resistant mutants of *Halobacterium* sp. NRC-1 showed up-regulation of the *rfa3–rfa8–ral* operon, consistent with the involvement of the corresponding RPA complex in haloarchaea in increased radiation resistance (DeVeaux et al., 2007). The *rfa3–rfa8–ral* operon was also shown to be upregulated after UV irradiation in *Halobacterium* sp. NRC-1 (McCready et al., 2005).

Another repair system in archaea which centers on the translesion DNA polymerases of the UmuC–DinB–Rad30–Rev1 superfamily A, is the functional equivalent of the mutagenic repair systems of bacteria and eukaryotes, and helps maintain genomic integrity under stressful conditions (Makarova et al., 2002). All sequenced haloarchaeal genomes encode an *umuC* gene homolog.

The exceptional ability of haloarchaea to tolerate UV light-induced damage likely also involves recombination activity which would facilitate recovery of stalled replication forks. Replication forks that have been abandoned by malfunctioning replisomes become prone to breakage (Kuzminov, 2001). RadA is the haloarchaeal homolog of RecA in bacteria and Rad51 in yeast, which catalyze strand invasion and exchange during homologous recombination. Two distinct *radA* genes are present in each of the sequenced haloarchaeal genomes. Deletion of *radA1* has been shown to cause severe UV sensitivity in *H. volcanii* (Woods and Dyall-Smith, 1997), and was shown to be upregulated after UV irradiation in *Halobacterium* sp. NRC-1 (McCready et al., 2005; Boubriak et al., 2008). The *radA2* gene, a second homolog of *recA* and *rad51*, encodes a protein with an unknown role in homologous recombination that was not induced after UV irradiation in *Halobacterium* sp. NRC-1 (McCready et al., 2005). In *E. coli, recJ* encodes a single-stranded DNA-specific exonuclease that is involved in the recovery of DNA replication at stalled forks, most likely by making DNA lesions at the stalled forks more accessible for repair (Courcelle et al., 2006). The RecJ-like exonuclease in archaea (Arj) contains a DnaJ-type zinc finger, and may be involved in recovery at stalled replication forks, similar

to its bacterial counterpart. *arj1* was upregulated after UV irradiation like *radA1* in *Halobacterium* sp. NRC-1 (McCready et al., 2005; Boubriak et al., 2008). The Holliday junction is a universal DNA intermediate in homologous recombination, and plays a crucial role in determination of genetic diversity and the repair of damaged chromosomes (Holliday, 1964; Liu and West, 2004). *hjr* codes for the haloarchaeal homolog of the Holliday junction resolvase, which functions as a dimer to introduce symmetric nicks to the Holliday junction in a metal-dependent manner (Nishino et al., 2001). A single copy of *recJ* and *hjr* and one of each of the two *arj* genes (*arj1* and *arj2*) are found in each of the ten sequenced haloarchaeal genera. The *rbl* genes encode predicted nucleases of the RecB family, which are associated with the RecBCD complex in *E. coli* and participate in the repair of double-stranded breaks by homologous recombination (Ren et al., 2007). *Halobacterium* sp. NRC-1 encodes two *rbl* genes, while the remaining haloarchaea encode a single gene.

6. Transcription and transcriptional regulators

When present, the eukaryotic homologs of the general transcriptional machinery are highly conserved across the sequenced haloarchaeal genomes (Table 3). The haloarchaea drive transcription using a single version of a eukaryotic RNA polymerase II-like enzyme encoded by 13 genes. The *rpoA′*, *A″*, *B′*, *B″*, and *H* genes are present in a cluster (in *H. turkmenica* and *N. pharaonis* the order is reversed), as are *rpoE′* and *E″*, and *rpoK* and *N*. The *rpoD*, *F*, *L*, and *P* genes are all present at separate loci in haloarchaea. Homologs of the eukaryotic general transcription factors TBP and TFIIB (TFB in archaea) associated with RNA polymerase are present in the haloarchaeal genomes. There are also two copies of the *tfs* gene and one copy of the *tfe* gene which are conserved in the haloarchaea. *tfs* codes for a transcription factor (TF) sometimes annotated as subunit M of RNA polymerase, and is a homolog of the eukaryotic TFIIS which induces cleavage of short mRNA molecules (Lange and Hausner, 2004). *tfe* codes for a homolog of the eukaryotic TFIIEα thought to enhance promoter strength (Grunberg et al., 2007). Genes homologous to the two helicase subunits of the TFIIH core complex (XPD and XPB) are conserved in haloarchaea and likely function in NER; however their role in transcription initiation has not been established. Additionally genes homologous to the remaining five TFIIH core subunits have not been identified in the sequenced haloarchaea. None of the genes coding for the other eukaryotic factors (TFIIEβ and TFIIF), the general cofactors, TATA-associated factors (TAFs), mediator complex, or upstream stimulatory activity (USA)-derived cofactors are present in any of the haloarchaea.

Multiple copies of genes coding for TBP, homologous to the eukaryotic TATA binding protein, and TFB, homologous to the eukaryotic TFIIB, are found in most of the haloarchaea (Figs. 3 and 4). *H. volcanii* has the largest total number of unique *tbp* and *tfb* genes, 15, with *H. lacusprofundi* and *Halobacterium* sp. R-1 having the next largest, 14. Interestingly, four of the haloarchaea contain a single copy of *tbp* genes, *H. marismortui*, *H. mukohataei*, *H. turkmenica* and *N. pharaonis*, but they all contain multiple *tfb* genes. The highest number of *tfb* gene paralogs (11) is found in *H. volcanii*, followed by *H. marismortui* and *H. lacusprofundi*, which have nine. The lowest number is found in *H. mukohataei*, which contains five paralogs. Many of the *tbp* and *tfb* genes are located on extrachromosomal elements, including five unique *tbp* genes and two *tfb* genes in *Halobacterium* sp. NRC-1, three *tbp* genes and three *tfb* genes in *H. lacusprofundi*, one *tbp* gene and five *tfb* genes in *H. volcanii*, three *tfb* genes in *H. marismortui*, and one *tfb* gene in *H. turkmenica*. Due to the large common regions on the pNRC and pHS plasmids in *Halobacterium* sp. NRC-1 and R-1, there are up to four identical copies of *tbp* genes on these extrachromosomal elements. This multiplicity of TBP and/or TFB factor genes has resulted in the hypothesis that a novel system for transcriptional regulation is present in the haloarchaea in which specific TBP–TFB pairs recognize distinct groups of promoters in response to specific conditions (Coker and DasSarma, 2007). Genes coding for termination/anti-termination factors homologous to NusA and NusG were also present in all the haloarchaea.

Although most of the basal transcription machinery in haloarchaea is homologous to the eukaryotic apparatus, transcriptional regulators in these organisms appear to be largely like those used in bacteria (Berquist et al., 2005). The largest family of predicted transcriptional regulators is the bacterial-type ArsR (HTH) family (*arl*R) (Table 3). This particular family of transcriptional regulators includes repressors that dissociate from DNA in the presence of metal ions to allow for transcription of the downstream genes (Wu and Rosen, 1993). In the metal ion rich environments that haloarchaea inhabit, this family of regulators likely plays a key role in their ability to thrive, and not surprisingly, is a major component of encoded transcriptional regulators. The sequenced haloarchaeal genomes contain up to 30 different *arl* genes, with the majority located on extrachromosomal elements. The conservation of this expanded family suggests the evolution of many analogous cellular functions among these regulators.

Another large family of transcriptional regulators is the PrlR family, which is named after the repressor PadA, a phenolic acid decarboxylase in *Pediococcus pentosaceus*. Some members of this family act as repressors by binding inverted repeats directly upstream of the transcription start site of target genes (Barthelmebs et al., 2000). Between two and 16 *prl* genes are present in the sequenced haloarchaeal genomes, and all but *H. mukohataei*, *N. pharaonis*, *H. walsbyi*, and *H. utahensis* have at least three on extrachromosomal elements.

A third family of regulators found in haloarchaea falls under the xenobiotic response element (Xre) HTH family. This type of transcriptional regulator, named *xrlR* in *Halobacterium* sp. NRC-1, was first identified in *Bacillus subtilis* as a repressor and has been subsequently identified in other bacteria and mold (Berquist et al., 2005). Three out of the four members of this family are present in *Halobacterium* sp. NRC-1 and are conserved across all ten sequenced haloarchaeal genera. The fourth member of this family (*xrlR2*) is conserved in five of the sequenced haloarchaeal genomes (*Halobacterium* sp. NRC-1, *H. volcanii*,

Table 3
Conserved genes involved in transcription and transcriptional regulation[a,b].

	COG	Number of homologs in NRC-1[c]	Number of homologs in Hma	Number of homologs in Nph	Number of homologs in Hwa	Number of homologs in Hla	Number of homologs in Hbo	Number of homologs in Hmu	Number of homologs in Hut	Number of homologs in Hvo	Number of homologs in Htu
rpoA′ & *A″*	COG0086: DNA-directed RNA polymerase, β' subunit/160 kD subunit	2	2	2	2	2	2	2	2	2	2
rpoB′ & *B″*	COG0085: DNA-directed RNA polymerase, β subunit/140 kD subunit	2	2	2	2	2	2	2	2	2	2
rpoD	COG0202: DNA-directed RNA polymerase, α subunit/40 kD subunit	1	1	1	1	1	1	1	1	1	1
rpoE′	COG1095: DNA-directed RNA polymerase, subunit E'	1	1	1	1	1	1	1	1	1	1
rpoE″	COG2093: DNA-directed RNA polymerase, subunit E"	1	1	1	1	1	1	1	1	1	1
rpoF	COG1460: Uncharacterized protein conserved in archaea	1	1	1	1	1	1	1	1	1	1
rpoH	COG2012: DNA-directed RNA polymerase, subunit H, RpoH/RPB5	1	1	1	1	1	1	1	1	1	1
rpoK	COG1758: DNA-directed RNA polymerase, subunit K/ω	1	1	1	1	1	1	1	1	1	1
rpoL	COG1761: DNA-directed RNA polymerase, subunit L	1	1	1	1	1	1	1	1	1	1
rpoN	COG1644: DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	1	1	1	1	1	1	1	1	1	1
rpoP	COG1996: DNA-directed RNA polymerase, subunit RPC10 (contains C4-type Zn-finger)	1	1	1	1	1	1	1	1	1	1
tfs	COG1594: DNA-directed RNA polymerase, subunit M/Transcription elongation factorTFIIS	2	2	2	2	2	2	2	2	2	2
tfeA	COG1675: Transcription initiation factor IIE, α subunit	1	1	1	1	1	1	1	1	1	1
tbp	COG2101: TATA-box binding protein (TBP), component of TFIID and TFIIIB	11 (6), 10	1	1	2	5, 3	3	1	2	4, 1	1
tfb	COG1405: Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB	7, 2	9, 3	7	8	9, 3	7	5	7	11, 5	6, 1
nusA	COG0195:Transcription elongation factor	1	1	1	1	1	1	1	1	1	1
nusG	COG0250: Transcription antiterminator	1	1	1	1	1	1	1	1	1	1
arlR associated with COG1733	COG1733: Predicted transcriptional regulators	1	2, 1	1	0	2, 1	3, 1	1	3	2, 1	2
arlR associated with COG3398	COG3398: Uncharacterized protein conserved in archaea	1	5, 3	3	1	2	2	2	1	2	3
arlR associated with COG1777	COG1777: Predicted transcriptional regulators	1	1	1	1	1	1	1	1	1	2
arlR associated with COG0640	COG0640: Predicted transcriptional regulators	13, 3	14, 4	10	7	11, 3	16, 3	14, 1	10	17, 4	18, 6
arlR associated with COG3355	COG3355: Predicted transcriptional regulator	3, 1	2, 1	1	0	2, 1	2, 1	1	1	2	1
arlR not associated with a COG	No associated COG	5 (2), 4	4, 1	2	6	2	6, 3	5, 1	6	5, 1	4, 2
prlR	COG1695: Predicted transcriptional regulators	12 (8), 7	13, 9	3	2	8, 4	7, 3	5	2	6, 3	16, 11
xrlR1	COG1992: Uncharacterized conserved protein	1	2	1	1	1	1	2	2	1	1
xrlR2	COG1476: Predicted transcriptional regulators	3 (2), 2	0	0	0	0	1, 1	1	2	2, 1	0

(*continued on next page*)

Table 3 (*continued*)

	COG	Number of homologs in NRC-1[c]	Number of homologs in Hma	Number of homologs in Nph	Number of homologs in Hwa	Number of homologs in Hla	Number of homologs in Hbo	Number of homologs in Hmu	Number of homologs in Hut	Number of homologs in Hvo	Number of homologs in Htu
xrlR3	COG1709: Predicted transcriptional regulator	1	1	1	1	1	1	1	1	1	1
xrlR4	COG1395: Predicted transcriptional regulator	1	1	1	1	1	1	1	1	1	1
aclR, asnC, cinR, nirD & *H*, & *trh3*	COG1522: Transcriptional regulators	11	11, 2	9, 2	8	9	12, 2	9	9	14, 1	11, 1
aclR not associated with a COG	No associated COG	1	1	1	1	1	1	1	1	1	1
calR	COG1497: Predicted transcriptional regulator	1	1	1	1	1	1	1	1	1	1
mbf1	COG1813: Predicted transcription factor, homolog of eukaryotic MBF1	1	1	1	1	1	1	1	1	1	1
phoU & *prp1*	COG0704: Phosphate uptake regulator	2	4	3, 1	3	3	2	3	3	2	3
rflR	COG1339: Transcriptional regulator of a riboflavin/FAD biosynthetic operon	1	1	1	1	1	1	1	1	1	1
sirR	COG1321: Mn-dependent transcriptional regulator	1	1	1	1	1	1	1	1	1	1
tzhR1	COG3357: Predicted transcriptional regulator containing an HTH domain fused to a Zn-ribbon	1	1	1	1	1	1	1	1	1	1
gul1	No associated COG	1	1	1	1	1	1	1	1	1	1
camR	COG3609: Predicted transcriptional regulators containing the CopG/Arc/MetJ DNA-binding domain	3, 1	2	2	1	2	2	2	1	3, 1	1
imd1	COG1475: Predicted transcriptional regulators	1	1	1	1	1	1	1	1	1	1
imd2	COG2524: Predicted transcriptional regulator, contains C-terminal CBS domains	1	1	1	1	1	1	1	1	1	1
sstR	COG1378: Predicted transcriptional regulators	3	6	1	2	6	5, 1	7	4	7, 1	5, 2
idr & *troR*	COG1321: Mn-dependent transcriptional regulator	3	2	1	1	1	2	3, 1	3	2	1
acrR	COG1309: Transcriptional regulator	2	7, 5	1	0	2	3, 2	3	2	3, 2	0
arcR	COG1414: Transcriptional regulator	1, 1	9, 5	3, 1	2	5, 2	7, 7	3, 1	0	18, 13	20, 14
glcK	COG1940: Transcriptional regulator/sugar kinase	1	1	0	1	1	1	1	1	1	1
marR	COG1733: Predicted transcriptional regulators	1	1	0	1	0	2, 2	2	0	0	1, 1
prtr1	COG4190: Predicted transcriptional regulator	3 (2), 2	1, 1	3	3	1, 1	0	0	1	2, 2	0
boa2-4 & *bolR*	COG3413: Predicted DNA binding protein	10 (7), 4	10	3	2	3	7, 3	2	3	12, 5	6, 2
bat, boa1, pcp	COG2202: FOG: PAS/PAC domain	3	4, 2	1	1	0	3, 2	3, 1	2	1	2
dmsR	COG3413: Predicted DNA binding protein	1	1	1	0	1	0	2, 1	0	1, 1	0

[a] When two numbers are expressed, the first number is the total number of genes, and the second number is the number of genes coded on extrachromosomal elements.

[b] Organism name abbreviations are as follows: *Halobacterium* sp. NRC-1 (NRC-1), *H. marismortui* (Hma), *N. pharaonis* (Nph), *H. walsbyi* (Hwa), *H. lacusprofundi* (Hla), *H. borinquense* (Hbo), *H. mukohataei* (Hmu), *H. utahensis* (Hut), *H. volcanii* (Hvo), and *H. turkmenica* (Htu).

[c] When three numbers are expressed, the first number is the total number of genes, the number in parentheses is the number of unique genes, and the third number is the number of genes coded on extrachromosomal elements.

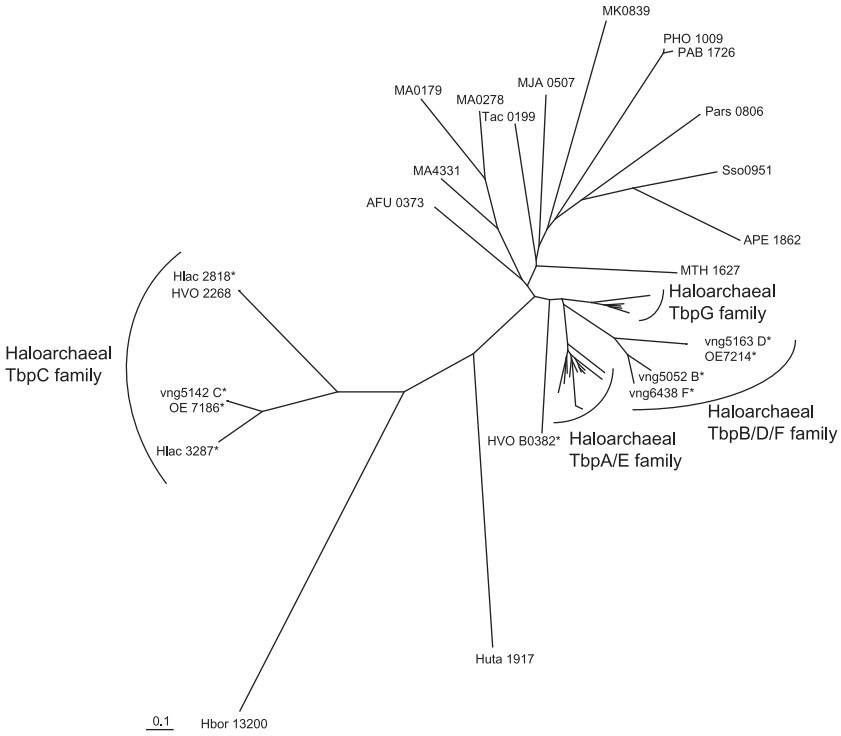

Fig. 3. Tbp Protein Clusters. Neighbor-joining radial tree of truncated haloarchaeal sequences and other archaeal sequences in COG2101 (*Archaeoglobus fulgidus*, *Aeropyrum pernix*, *Methanosarcina acetivorans*, *Methanococcus jannaschii*, *Methanopyrus kandleri*, *Methanothermobacter thermautotrophicus*, *Pyrococcus abyssi*, *Pyrobaculum arsenaticum*, *Pyrococcus horikoshii*, *Sulfolobus solfataricus*, *Thermoplasma acidophilum*). *indicates proteins coded on extrachromosomal elements. Haloarchaeal TbpA/E family includes VNG2243, VNG5039*, rrnAC0681, HVO_0158, Hbor_00070, Hmuk_0721, Huta_1250, Hlac_2629, NP1064A, HQ3410A, Htur_0133, OE4146, and OE7045*. Haloarchaeal TbpG family includes HVO_1727, Hbor_14680, Hlac_1523, Hlac_3413*, and HQ2527A.

H. borinquense, *H. mukohataei*, and *H. utahensis*). Two of the *xrlR2* genes in *Halobacterium* sp. NRC-1 are encoded on the common region of pNRC100 and pNRC200, and one of the *H. volcanii* and the *H. borinquense xrlR2* genes are located on the pHV4 and pHB400 extrachromosomal elements, respectively.

Other conserved families of transcriptional regulators include the *aclR*, *camR*, *calR*, *cinR*, *imd1* and *2*, *gul1*, *mbf1*, *phoU*, *prp1*, *rflR*, *sirR*, *sstR*, *thr3*, *troR*, and *tzhR1* genes (Table 3). The AsnC, CinR, NirD and H, and Trh3 proteins and many of the AclR proteins are associated with a single transcriptional regulator family (COG1522), and at least one gene coding for these proteins is found on extrachromosomal elements in *H. marismortui*, *H. volcanii*, *H. borinquense*, *N. pharaonis*, and *H. turkmenica*. Similarly, Ird and TroR are both associated with the manganese-dependent transcriptional regulator family (COG1321), and one of the two *H. mukohataei ird* genes is located on pHM61. *sirR* codes for a manganese-dependent transcriptional regulator and is present as a single copy in all ten haloarchaeal genera. *phoU* codes for a homolog of the repressor of the phosphate ABC transporter (Liu et al., 2005) and is associated with the same phosphate uptake regulator family as *prp1* (COG0704). At least one copy of each gene is present in all of the sequenced haloarchaeal genomes with multiple copies in seven of the genomes. One copy of the *phoU* gene in *N. pharaonis* is present on the extrachromosomal element pNP131. *tzhR1* codes for a transcriptional regulator with an HTH domain fused to a Zn-ribbon and is present as a single copy conserved in all ten genomes. *gul1* is conserved among all sequenced haloarchaeal genomes and contains a repressor HTH domain. Several other families of transcriptional regulators include genes coded on extrachromosomal elements, such as one of the *camR* genes coded on pNRC200 in *Halobacterium* sp. NRC-1 and pHV1 in *H. volcanii*, one of the *sstR* genes coded on pHV4 in *H. volcanii* and pHB200 in *H. borinquense*, and two of the *sstR* genes coded on pHT85 in *H. turkmenica*.

The bacterio-opsin activator protein *bat* gene of *Halobacterium* sp. NRC-1 is a prototype of the Bat transcriptional activator family. It contains both sensor domains (PAS/PAC and GAF motifs), as well as the specific pfam HTH10 family DNA-binding domain (Betlach et al., 1989; DasSarma, 2004), and proteins containing these three domains are found in all ten haloarchaeal genera. In *Halobacterium* sp. NRC-1, the Bat protein is responsible for coordinating regulation of bacteriorhodopsin (the purple membrane protein, containing bacterio-opsin protein + retinal cofactor) and this gene is conserved in seven of the ten sequenced genera (absent in *H. borinquense*, *H. lacusprofundi*, and *H. walsbyi*). The bacterio-opsin and retinal biosynthetic genes are conserved in seven of the ten genomes (bacterio-opsin genes are absent in *H. volcanii*, *H. borinquense*, and *H. turkmenica*, although

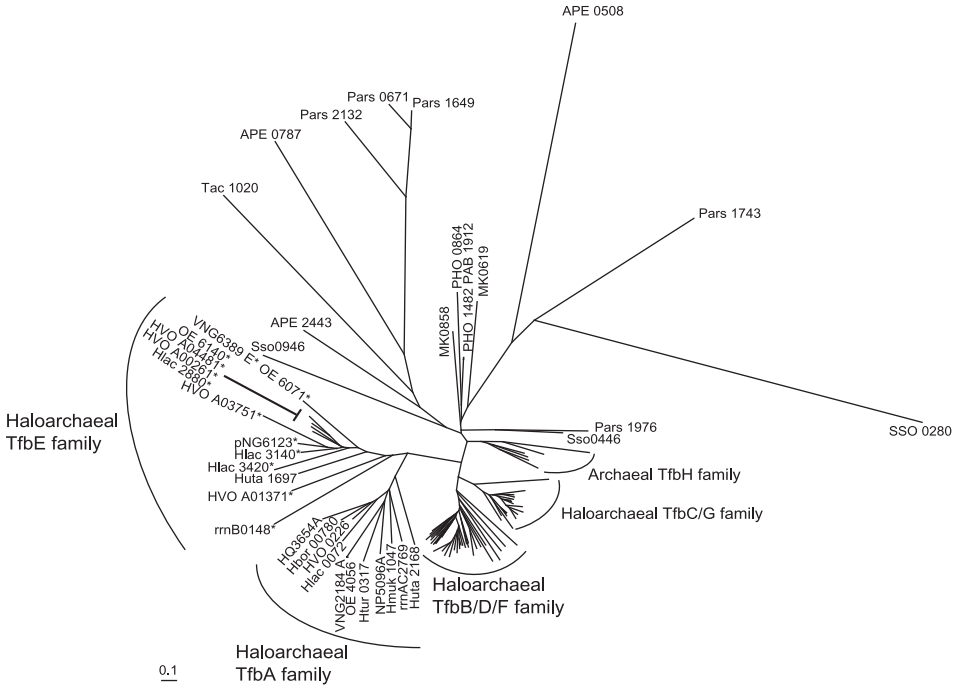

Fig. 4. Tfb Protein Clusters. Neighbor-joining radial tree of truncated haloarchaeal sequences and other archaeal sequences in COG1405 (*Archaeoglobus fulgidus*, *Aeropyrum pernix*, *Methanosarcina acetivorans*, *Methanococcus jannaschii*, *Methanopyrus kandleri*, *Methanothermobacter thermautotrophicus*, *Pyrococcus abyssi*, *Pyrobaculum arsenaticum*, *Pyrococcus horikoshii*, *Sulfolobus solfataricus*, and *Thermoplasma acidophilum*). *indicates proteins coded on extrachromosomal elements. Haloarchaeal TfbB/D/F family includes VNG0315, VNG0734, VNG0869, rrnAC1126, rrnAC1357, rrnAC1875, rrnAC2051, pNG6072*, HVO_0795, HVO_1052, HVO_1676, HVO_B0285*, Hbor_15380, Hbor_22340, Hbor_23010, Hbor_24220, Hmuk_1991, Hmuk_2412, Hmuk_3030, Huta_0083, Huta_0554, Huta_1187, Huta_1896, Hlac_0601, Hlac_1327, Hlac_1513, NP2220A, NP2246A, NP4326A, HQ1153A, HQ1625A, HQ1689A, HQ2571A, HQ3408A, Htur_0836, Htur_2440, OE1478, OE2084, and OE2281. Haloarchaeal TfbC/G family includes VNG0254, VNG6351*, rrnAC0016, HVO_0733, HVO_1478, Hbor_17550, Hbor_24820, Hmuk_2679, Huta_1219, Hlac_0309, Hlac_1495, NP1684A, NP4822A, NP5186A, HQ1227A, HQ1867A, Htur_0478, Htur_2586, Htur_5247*, OE1399, and OE6026*. Archaeal TfbH family includes AFU1299, MA0610, MJA0782, MTH0885, Tac0940, and Tac0945.

retinal biosynthetic genes are present). Several bacterio-opsin like activator genes (*boa* and *bolR*) are found in each of the ten genomes. Bat, Boa1, and Pcp (PAS/PAC domain containing protein) proteins are all within a single protein family (COG2202). The three remaining Boa proteins (Boa2, 3, and 4) and the four BolR (bacterio-opsin like activator) proteins (BolR1, R2, R3, and R4) of *Halobacterium* sp. NRC-1 (members of COG3413) form an expanded family with the homologs from all the other nine haloarchaeal genera.

Several families of transcriptional regulator genes (typified by *acrR*, *arcR*, *dmsR*, *glcK*, *marR*, and *prtR1*) are conserved in a fraction of the sequenced haloarchaeal genomes, and some are encoded on extrachromosomal elements (Table 3). The *dmsR* gene, located on the *Halobacterium* sp. NRC-1 chromosome, has been studied using transcriptomic and genetic approaches. The *dms* operon contains five transcriptionally linked genes (*dmsEABCD*) coding for a DMSO/TMAO reductase and molecular chaperone, and *Halobacterium* sp. NRC-1 was shown to be able to utilize both DMSO and TMAO as terminal electron acceptors and grow under anaerobic conditions (Müller and DasSarma, 2005). Deletion of the *dmsR* gene, located immediately upstream of the transcription unit, resulted in decreased survival under anaerobic conditions and decreased transcription of the *dms* operon, indicating the regulator is a transcriptional activator. *H. marismortui*, *H. mukohataei*, and *H. volcanii* have homologs of the DMSO/TMAO reductase and *dmsR* regulator, suggesting that each is able to utilize these compounds in anaerobic respiration, while *H. lacusprofundi* and *N. pharanosis* only have a regulatory gene similar to *dmsR*.

7. Translation

Translation, the process of mRNA-encoded protein synthesis, requires a complex apparatus composed of the ribosome, tRNAs and additional protein factors, including amino-acyl tRNA sythetases, and constitutes the largest number of conserved genes in the haloarchaea with most present on the chromosomes (Table 4). Nearly all eukaryotic translation initiation factors, which are known to bind to the small ribosomal subunit before association with mRNA, have archaeal homologs (Marintchev and Wagner, 2004). Out of eight eukaryotic translation initiation factor (*eif*) genes, seven are highly conserved in haloarchaea. In eukaryotes, eIF1 and eIF1A are important for scanning and start codon selection and can discriminate against non-AUG codons or AUG codons located near the 5′-end. During start codon recognition these factors act in an antagonistic manner, where eIF1 inhibits premature initiation at non-canonical and alternative AUG start codons, and eIF1A promotes initiation (Pestova et al., 1998; Mitchell and Lorsch, 2008). Initially based on sequence similarity,

Table 4
Conserved genes involved in translation[a,b].

	COG	Number of homologs in NRC-1	Number of homologs in Hma	Number of homologs in Nph	Number of homologs in Hwa	Number of homologs in Hla	Number of homologs in Hbo	Number of homologs in Hmu	Number of homologs in Hut	Number of homologs in Hvo	Number of homologs in Htu
eif1a	COG0361: Translation initiation factor 1 (IF-1)	2	2	2	2	2	2	2	2	2, 1	2
sui1	COG0023: Translation initiation factor 1 (eIF-1/SUI1) and related proteins	1	1	1	1	1	1	1	1	1	1
eif2a	COG1093: Translation initiation factor 2, α subunit (eIF-2α)	1	1	1	1	1	1	1	1	1	1
eif2b	COG1601: Translation initiation factor 2, β subunit (eIF-2β)/eIF-5*N*-terminal domain	1	2	2	1	1	2	2	2	2	1
eif2ba & *2bd*	COG1184: Translation initiation factor 2B subunit, eIF-2B α/β/δ family	2	3	2	1	2	3	2	1	3	3
eif2g	COG0050: GTPases - translation elongation factors	1	1	1	1	1	1	1	1	1	1
eif5a	COG0231: Translation elongation factor P (EF-P)/ translation initiation factor 5A (eIF-5A)	1	1	1	1	1	1	1	1	1	1
eif6	COG1976: Translation initiation factor 6 (eIF-6)	1[c]	1[d]	1	1	1	1	1	1	1	1
dys	COG1899: Deoxyhypusine synthase	1	1	1	1	1	2, 1	1	1	2, 1	1
fib	COG1889: Fibrillarin-like rRNA methylase	1	1	1	1	1	1	1	1	1	1
infB	COG0532: Translation initiation factor 2 (IF-2; GTPase)	1	1	1	1	1	1	1	1	1	1
eef1a	COG5256: Translation elongation factor EF-1α (GTPase)	1	1	1	1	1	2	1	1	2	1
eef1b	COG2092: Translation elongation factor EF-1β	1	1	1	1	1	1	1	1	1	1
eef2	COG0480: Translation elongation factors (GTPases)	1	1	1	1	1	1	1	1	1	1
alaS	COG0013: Alanyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
argS	COG0018: Arginyl-tRNA synthetase	1, 1	1	1	1	1	1	1	1	1	1
aspS	COG0017: Aspartyl/asparaginyl-tRNA synthetases	1	1	1	1	1	1	1	1	1	1
cysS	COG0215: Cysteinyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
gltS	COG0008: Glutamyl- and glutaminyl-tRNA synthetases	1	1	1	1	1	1	1	1	1	1
glyS	COG0423: Glycyl-tRNA synthetase (class II)	1	1	1	1	1	1	1	1	1	1
hisS	COG0124: Histidyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
ileS	COG0060: Isoleucyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
leuS	COG0495: Leucyl-tRNA synthetase	1	2	1	1	1	1	2	2	1	1
lysS	COG1384: Lysyl-tRNA synthetase (class I)	1	1	1	1	1	1	1	1	1	1
metS	COG0143: Methionyl-tRNA synthetase & COG0073: EMAP domain	1	1	1	1	1	1	1	1	1	1
pheS	COG0016: Phenylalanyl-tRNA synthetase α subunit	1	1	1	1	1	1	1	1	1	1
pheY	COG0072: Phenylalanyl-tRNA synthetase β subunit	1	1	1	1	1	1	1	1	1	1
proS	COG0442: Prolyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
serS	COG0172: Seryl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
thrS	COG0441: Threonyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
trpS	COG0180: Tryptophanyl-tRNA synthetase	2	1	1	1	1	2	1	1	2, 1	2
tyrS	COG0162: Tyrosyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
valS	COG0525: Valyl-tRNA synthetase	1	1	1	1	1	1	1	1	1	1
gatA	COG0154: Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	1	1	1	1	1	1	1	1	1	1
gatB1	COG2511: Archaeal Glu-tRNAGln amidotransferase subunit E (contains GAD domain)	1	1	1	1	1	1	1	1	1	1
gatB2	COG0064: Asp-tRNAAsn/Glu-tRNAGln amidotransferase B subunit (PET112 homolog)	1	1	1	1	1	1	1	1	1	1
gatC	COG0721: Asp-tRNAAsn/Glu-tRNAGln amidotransferase C subunit	1	1	1	1	1	1	1	1	1	1

(*continued on next page*)

Table 4 (*continued*)

	COG	Number of homologs in NRC-1	Number of homologs in Hma	Number of homologs in Nph	Number of homologs in Hwa	Number of homologs in Hla	Number of homologs in Hbo	Number of homologs in Hmu	Number of homologs in Hut	Number of homologs in Hvo	Number of homologs in Htu
rpl10e	COG0197: Ribosomal protein L16/L10E	1	1	1	1	1	1	1	1	1	1
rpl10p	COG0244: Ribosomal protein L10	1	1	1	1	1	1	1	1	1	1
rpl11p	COG0080: Ribosomal protein L11	1	1	1	1	1	1	1	1	1	1
rpl12p	COG2058: Ribosomal protein L12E/L44/L45/RPP1/RPP2	1	1	1	1	1	1	1	1	1	1
rpl13p	COG0102: Ribosomal protein L13	1	1	1	1	1	1	1	1	1	1
rpl14p	COG0093: Ribosomal protein L14	1	1	1	1	1	1	1	1	1	1
rpl15e	COG1632: Ribosomal protein L15E	1	1	1	1	1	1	1	1	1	1
rpl15p	COG0200: Ribosomal protein L15	1	1	1	1	1	1	1	1	1	1
rpl18e	COG1727: Ribosomal protein L18E	1	1	1	1	1	1	1	1	1	1
rpl18p	COG0256: Ribosomal protein L18	1	1	1	1	1	1	1	1	1	1
rpl19e	COG2147: Ribosomal protein L19E	1	1	1	1	1	1	1	1	1	1
rpl1p	COG0081: Ribosomal protein L1	1	1	1	1	1	1	1	1	1	1
rpl21e	COG2139: Ribosomal protein L21E	1	1	1	1	1	1	1	1	1	1
rpl22p	COG0091: Ribosomal protein L22	1	1	1	1	1	1	1	1	1	1
rpl23p	COG0089: Ribosomal protein L23	1	1	1	1	1	1	1	1	1	1
rpl24e	COG2075: Ribosomal protein L24E	1	1	1	1	1	1	1	1	1	1
rpl24p	COG0198: Ribosomal protein L24	1	1	1	1	1	1	1	1	1	1
rpl29p	COG0255: Ribosomal protein L29	1	1	1	1	1	1	1	1	1	1
rpl2p	COG0090: Ribosomal protein L2	1	1	1	1	1	1	1	1	1	1
rpl30p	COG1841: Ribosomal protein L30/L7E	1	1	1	1	1	1	1	1	1	1
rpl31e	COG2097: Ribosomal protein L31E	1	1	1	1	1	1	1	1	1	1
rpl32e	COG1717: Ribosomal protein L32E	1	1	1	1	1	1	1	1	1	1
rpl37e	COG2126: Ribosomal protein L37E	1	0	1	1	1	1	0	1	1	1
rpl39e	COG2167: Ribosomal protein L39E	1	1	1	1	1	1	1	1	1	1
rpl3p	COG0087: Ribosomal protein L3	1	1	1	1	1	1	1	1	1	1
rpl40e	COG1552: Ribosomal protein L40E	1	1	1	1	1	0	0	0	1	0
rpl44e	COG1631: Ribosomal protein L44E	1	1	1	1	1	1	1	1	1	1
rpl4e	COG0088: Ribosomal protein L4	1	1	1	1	1	1	1	1	1	1
rpl5p	COG0094: Ribosomal protein L5	1	1	1	1	1	1	1	1	1	1
rpl6p	COG0097: Ribosomal protein L6P/L9E	1	1	1	1	1	1	1	1	1	1
rps10p	COG0051: Ribosomal protein S10	1	1	1	1	1	1	1	1	1	1
rps11p	COG0100: Ribosomal protein S11	1	1	1	1	1	1	1	1	1	1
rps12p	COG0048: Ribosomal protein S12	1	1	1	1	1	1	1	1	1	1
rps13p	COG0099: Ribosomal protein S13	1	1	1	1	1	1	1	1	1	1
rps14p	COG0199: Ribosomal protein S14	1	1[d]	2	1	1	1	1	1	1	1
rps15p	COG0184: Ribosomal protein S15P/S13E	1	1	1	1	1	1	1	1	1	1
rps17e	COG1383: Ribosomal protein S17E	1	1	1	1	1	1	1	1	1	1
rps17p	COG0186: Ribosomal protein S17	1	1	1	1	1	1	1	1	1	1
rps19e	COG2238: Ribosomal protein S19E (S16A)	1	1	1	1	1	1	1	1	1	1
rps19p	COG0185: Ribosomal protein S19	1	1	1	1	1	1	1	1	1	1
rps24e	COG2004: Ribosomal protein S24E	1	1	1	1	1	1	1	1	1	1
rps27ae	COG1998: Ribosomal protein S27AE	1	1	1	0	0	0	1	1	0	1
rps27e	COG2051: Ribosomal protein S27E	1	1	1	1	1	1	1	1	1	1
rps28e	COG2053: Ribosomal protein S28E/S33	1	1[d]	1	1	1	1	1	1	1	1
rps2p	COG0052: Ribosomal protein S2	1	1	1	1	1	1	1	1	1	1

rps3e	COG1890: Ribosomal protein S3AE	1	1	1	1	1	1	1	1	1	1
rps3p	COG0092: Ribosomal protein S3	1	1	1	1	1	1	1	1	1	1
rps4e	COG1471: Ribosomal protein S4E	1	1	1	1	1	1	1	1	1	1
rps4p	COG0522: Ribosomal protein S4 and related proteins	1	1	1	1	1	1	1	1	1	1
rps5p	COG0098: Ribosomal protein S5	1	1	1	1	1	1	1	1	1	1
rps6e	COG2125: Ribosomal protein S6E (S10)	1	1	1	1	1	1	1	1	1	1
rps7p	COG0049: Ribosomal protein S7	1	1	1	1	1	1	1	1	1	1
rps8e	COG2007: Ribosomal protein S8E	1	1	1	1	1	1	1	1	1	1
rps8p	COG0096: Ribosomal protein S8	1	1	1	1	1	1	1	1	1	1
rps9p	no associated COG	1	1	1	1	1	1	1	1	1	1
rphs6	COG1358: Ribosomal protein HS6-type (S12/L30/L7a)	1	1	1	1	1	1	1	1	1	1
drg	COG1163: Predicted GTPase	1	1	1	1	1	1	1	1	1	1
hemA	COG0373: Glutamyl-tRNA reductase	1	1	1	1	1	1	1	1	1	1
hemK	COG2890: Methylase of polypeptide chain release factors	1	1	1	1	1	1	1	1	1	1
nop58	COG1498: Protein implicated in ribosomal biogenesis, Nop56p homolog	1	1	1	1	1	1	1	1	1	1
pimT1	COG2518: Protein-L-isoaspartate carboxylmethyltransferase	1	1	1	1	1	1	1	1	1	1
pimT2	COG2519: tRNA (1-methyladenosine) methyltransferase and related methyltransferases	1	1	1	1	1	1	1	1	1	1
cca	COG1746: tRNA nucleotidyltransferase (CCA-adding enzyme)	1	1	1	1	1	1	1	1	1	1
endA	COG1676: tRNA splicing endonuclease	1	1	1	1	1	1	1	1	1	1
sua	COG0009: Putative translation factor (SUA5)	1	1	1	1	1	1	1	1	1	1
truA	COG0101: Pseudouridylate synthase	1	1	1	1	1	1	1	1	1	1
cna	COG0144: tRNA and rRNA cytosine-C5-methylases	1	1	1	1	1	1	1	1	1	1
map	COG0024: Methionine aminopeptidase	1	1	1	1	1	1	1	1	1	1
rimI	COG0456: Acetyltransferases	1	1	1	1	1	1	1	1	1	1
rimK	COG0189: Glutathione synthase/Ribosomal protein S6 modification enzyme (glutaminyl transferase)	1	1	1	1	1	1	1	1	1	1
tgtA1	COG0343: Queuine/archaeosine tRNA-ribosyltransferase	1	1	1	0	1	1	1	1	1	1
tgtA2	COG1549: Queuine tRNA-ribosyltransferases, contain PUA domain	1	1	1	0	1	1	1	1	1	1
trm1	COG1867: N2,N2-dimethylguanosine tRNA methyltransferase	1	1	1	1	1	1	1	1	1	1
gbp5	COG0012: Predicted GTPase, probable translation factor	1	1	1	1	1	1	1	1	1	1
rne	COG1530: Ribonucleases G and E	1	1	1	1	1	1	1	1	1	1
lds	COG1798: Diphthamide biosynthesis methyltransferase	1	1	1	1	1	1	1	1	1	1
ftsJ	COG0293: 23S rRNA methylase	1	1	1	1	1	1	1	1	1	1
spoU	COG0565: rRNA methylase	1	1	1	1	1	1	1	1	1	1
erf1	COG1503: Peptide chain release factor 1 (eRF1)	2	2	2	2	2	2	2	2	2	3
5S rRNA		1	4, 2	1	2	3, 1	2	3, 1	2	2	4
16S rRNA		1	3, 1	1	2	3, 1	2	3, 1	1	2	3
23S rRNA		1	3, 1	1	2	3, 1	2	3, 1	1	2	3

[a] When two numbers are expressed, the first number is the total number of genes, and the second number is the number of genes coded on extrachromosomal elements.

[b] Organism name abbreviations are as follows: *Halobacterium* sp. NRC-1 (NRC-1), *H. marismortui* (Hma), *N. pharaonis* (Nph), *H. walsbyi* (Hwa), *H. lacusprofundi* (Hla), *H. borinquense* (Hbo), *H. mukohataei* (Hmu), *H. utahensis* (Hut), *H. volcanii* (Hvo), and *H. turkmenica* (Htu).

[c] The original annotation of *Halobacterium* sp. NRC-1 did not contain *eif6*, this gene was later identified (unpublished data).

[d] The original annotation of *H. marismortui* did not contain *eif6*, *rps14p*, or *rps27ae*, these genes were identified in the archaeal COG study (Makarova et al., 2007).

only homologs to eIF1A were identified in archaea; however, a functional homolog to eIF1 has recently been identified and studied in *S. solfataricus* (Hasenöhrl et al., 2006; Hasenöhrl et al., 2009). Both eIF1 and eIFA, individually and together, stimulated eIF2 binding to the 30S ribosome in *S. solfataricus*, and the eIF1 homolog was shown to inhibit binding of the 30S subunit to non-canonical start codons (Hasenöhrl et al., 2006; Hasenöhrl et al., 2009). Two distinct *eif1a* homologs and one *eif1* (*sui1*) homolog are present in each of the sequenced haloarchaeal genomes, and one of the *H. volcanii eif1a* genes is coded on the extrachromosomal element pHV4. eIF2 is a heterotrimer, comprised of α, β, and γ subunits, and functions in selection and recruitment of Met-tRNA to the 40S ribosomal subunit and also controls start site recognition. eIF2 requires eIF5 to act as a GTPase-activator protein (GAP) by stabilizing a transition state in GTP hydrolysis in order to function (Marintchev and Wagner, 2004). eIF2ba is the only translation initiation factor that is not fully conserved in haloarchaea (absent in *H. marismortui*, *H. utahensis*, and *H. walsbyi*), and in eukaryotes, it is involved in inhibition by eIF2a phosphorylation, where it was not found to be essential (Marintchev and Wagner, 2004). eIF2ba and eIF2bd are both associated with COG1184, and the gene coding for eIF2bd is present in all sequenced haloarchaeal genomes, with *H. marismortui*, *H. volcanii*, *H. borinquense*, and *H. turkmenica* harboring multiple copies. Genes coding eIF6 have been identified in *Halobacterium* sp. NRC-1 and *H. marismortui* (Makarova et al., 2007). Eukaryotic eIF6 prevents 60S and 40S ribosomal subunit interactions by binding specifically to free 60S ribosomal subunits (Si and Maitra, 1999). Similarly, the *S. solfataricus* eIF6 homolog binds specifically to the 50S ribosomal subunit preventing the 70S ribosome formation (Benelli et al., 2009). Conserved among all haloarchaeal genomes are the *dys* and *fib* genes, which encode the enzyme deoxyhypusine synthase (DHS) that catalyzes a reaction in which the initiation factor 5A undergoes an unusual post-translation modification whereby a specific conserved lysine residue is transformed into the amino acid hypusine (Brochier et al., 2004). *H. volcanii* and *H. borinquense* both code a second *dys* gene on the pHV3 and pHB500 extrachromosomal elements, respectively.

Among translational elongation factors, eEF1a has been shown to be an indispensable GTPase involved in the translation process in eukaryotes (Andersen et al., 2001). The *eef1a* gene is conserved in the sequenced haloarchaeal genomes, with multiple copies of the gene found in both *H. volcanii* and *H. borinquense*. Genes coding for two additional elongation factors, eEF1b and eEF2, are also present in the genomes of all ten haloarchaeal genera.

Nineteen amino-acyl tRNA synthetases are conserved in all ten haloarchaeal genomes. As observed in some other organisms (Diaz-Lazcoz et al., 1998), seven haloarchaea contain multiple copies of individual tRNA synthetases. *H. marismortui*, *H. mukohataei*, and *H. utahensis* each contain two copies of leucyl-tRNA synthetase and *Halobacterium* sp. NRC-1, *H. volcanii*, *H. borinquense*, and *H. turkmenica* each contain two copies of tryptophanyl-tRNA synthetase, with one *H. volcanii* gene coded on pHV4. The *Halobacterium* sp. NRC-1 arginyl-tRNA synthetase gene, *argS*, is found on pNRC200, and is likely to result from lateral gene transfer (Kennedy, 2003). Of the two different tRNA-dependent amidotransferases (AdTs) known, only GatCAB is present in most bacteria and archaea, including all the haloarchaea. These enzymes are essential because they are required for the proper charging of glutaminyl-tRNA (Gln-tRNA) with its conjugate amino acid (Sheppard and Söll, 2008). The *gatA, gatB,* and *gatC* genes are conserved in all ten sequenced haloarchaeal genera.

All of the sequenced haloarchaeal genomes have at least one rRNA region coded on the chromosome located in regions with shared synteny among the genomes (Fig. 5). Following the 16S rRNA gene, Ala-tRNA, 23S rRNA, 5S rRNA, and Cys-tRNA genes are coded; however, *H. marismortui* and *H. utahensis*, each have an additional hypothetical open reading frame inserted between the 5S rRNA and Cys-tRNA genes, and *H. utahensis* also has a second 5S rRNA gene inserted. *H. marismortui*, *H. volcanii*, *H. mukohataei*, *H. borinquense*, *H. lacusprofundi*, and *H. walsbyi* all have an additional rRNA region coded on the chromosome. *H. turkmenica* has three rRNA regions on the chromosome and a lone fourth 5S rRNA gene. The second chromosomally coded rRNA region in the *H. borinquense*, *H. lacusprofundi*, *H. volcanii*, and *H. walsbyi* genomes, as well as two of the *H. turkmenica* rRNA regions, are syntenic. In addition to the chromosomal rRNA regions, *H. marismortui*, *H. mukohataei*, and *H. lacusprofundi* have rRNA regions on extrachromosomal elements. The presence of these additional complete rRNA regions led to the *H. marismortui* and *H. lacusprofundi* extrachromosomal elements being designated as the second chromosomes (chromosome II), while the *H. mukohataei* extrachromosomal element is designated pHM61. *H. marismortui* has an additional 5S rRNA gene coded on the pNG700 extrachromosomal element.

Interestingly, the multiple 16S rRNA genes of some haloarchaea are highly divergent in the same organism. For example, the 16S rRNA genes on the *H. marismortui* chromosome differ by 5.7% and the third extrachromosomal gene differs by 5.2% from one chromosomal copy but only 0.8% from the other. Likewise, the extrachromosomal 16S rRNA gene on pHM61 in *H. mukohataei* is identical to one chromosomal copy, and these genes are highly divergent (9.3%) from the other chromosomal gene.

Ribosomal protein (r-protein) genes are clustered in many microorganisms. For example, in *E. coli*, 32 r-proteins from both subunits and two translation-related proteins are grouped into seven well-studied operons: α, *L10*, *L11*, *S10*, *S20*, *str* and *spc* (Nomura et al., 1984). In archaea, genes coding the large and small ribosomal subunits are also highly clustered, although they are not syntenic across the domain. The 50S subunit consists of 30 different proteins, all of which are highly conserved in the haloarchaea, except for L37E and L40E. The gene coding for the L37e protein is absent in a number of other archaeal genomes, including *P. aerophilum*. The 30S subunit has 25 different proteins which are also highly conserved among all ten haloarchaeal genera except for S27AE, which is missing in *H. volcanii*, *H. borinquense*, *H. lacusprofundi*, and *H. walsbyi*. The original annotation of *H. marismortui* did not contain a S14P or S28E homolog, but these have

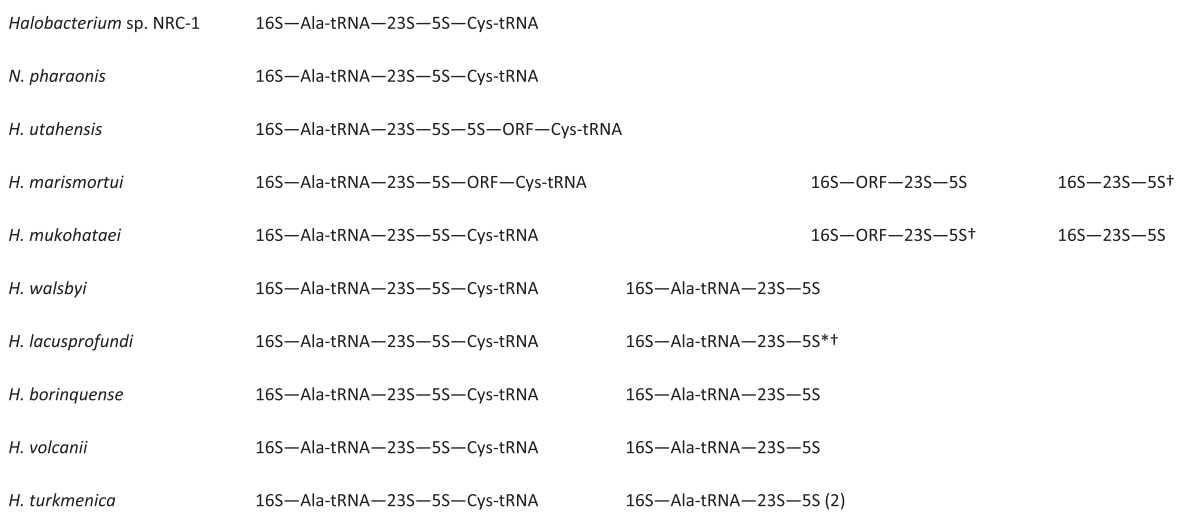

Fig. 5. Schematic representation of the rRNA regions coded on the chromosome and extrachromosomal elements of haloarchaea. Three rRNA regions are present on extrachromosomal elements indicated with † and one is similar to a chromosomal rRNA region indicated with*. Two nearly identical *H. turkmenica* rRNA regions are indicated with (2). In addition to the rRNA genes illustrated, not shown are lone fourth 5S rRNA genes, identified on the chromosome of *H. turkmenica* and pNG700 of *H. marismortui*. Different open reading frames (labeled ORF) are present within rRNA gene regions of *H. utahensis*, *H. marismortui*, *H. mukohataei*, the functions of which are unknown.

recently been identified in a study of conserved archaeal proteins (Makarova et al., 2007). The remaining non-conserved ribosomal proteins are primarily associated with regions of the rRNA thought to be additions to the ribosomal machinery, likely coming into existence at a later time in evolution (Wang et al., 2009).

Developmentally regulated G-proteins (DRGs) are a highly conserved family of GTP-binding proteins found in archaea, plants, fungi, and animals, indicating their important roles in fundamental pathways. The ability of DRGs to bind nucleotide substrates without assistance, their slow rate of GTP hydrolysis, heat stress activation, and domain conservation suggest a possible role in ribosome assembly in response to stress (O'Connell et al., 2009). The *drg* gene present in *Halobacterium* sp. NRC-1 is conserved in all of the other nine sequenced haloarchaeal genera. The initial reaction of tetrapyrrole formation in archaea is catalyzed by a NADPH-dependent glutamyl-tRNA reductase (GluTR) (Moser et al., 1999). The *hem*A gene encodes this GluTR and is conserved among all ten haloarchaeal genera. The eukaryotic nucleolus contains a diverse population of small nucleolar RNAs (snoRNAs) essential for ribosome biogenesis. The box C/D snoRNA family possesses conserved nucleotide boxes C and D that are multifunctional elements required for snoRNA processing, snoRNA transport to the nucleolus, and 2′-O-methylation of ribosomal RNA. Nop58 is a core protein of the box C/D snoRNP complex and the gene encoding this protein is found in all the sequenced haloarchaeal genomes.

Several other miscellaneous translation genes are conserved among all ten sequenced haloarchaeal genera, including L-isoaspartyl protein carboxyl methyltransferase (*pimT*), tRNA nucleotidyltransferase (*cca*), tRNA intron endonuclease (*endA*), predicted translation factor (*sua*), pseudouridylate synthase I (*truA*), methionine aminopeptidase (*map*), rRNA methyltransferase (*ftsJ* and *spoU*), several genes associated with putative translational functions, including the predicted GTPase, *gdp5*, and 13 additional genes of unknown function associated with translational protein clusters.

Accurate translation termination is essential for cell viability. In eukaryotes, this process is strictly maintained by two proteins, eukaryotic release factor 1 (eRF1), which recognizes all stop codons and hydrolyzes peptidyl-tRNA bonds, and eukaryotic release factor 3 (eRF3), a GTPase which enhances eRF1 activity (Inagaki and Doolittle, 2000). All ten sequenced haloarchaeal genera possess genes coding for two–three homologs of eRF1, but lack C-terminal tails for eRF3 binding. Consistent with this observation, an *erf3* homolog was not identified in any of the sequenced haloarchaeal genomes.

8. Concluding remarks

The information transfer system of haloarchaea, as in other archaea, has both bacterial and eukaryotic features. While genes coding for the biosynthesis of macromolecules are homologous to eukaryotic genes, the genome organization and regulators of gene expression are bacterial-like. Proteins required for these fundamental processes are highly conserved and often encoded in syntenic regions among the sequenced members of the ten haloarchaeal genera. With the exception of only a few, these critical proteins are chromosomally encoded, suggesting that the large extrachromosomal elements or plasmids, while common, do not encode these essential functions. Some exceptions such as the *argS* gene of *Halobacterium* sp. NRC-1 encoded on pNRC200 are likely to be indicative of genes acquired by lateral gene transfer.

A number of information transfer genes form expanded families (*e.g.*, *orc1*/*cdc6*, *act*, *tbp*, *tfb*, *arl*, *prl*, *xrl*) and some members of these families are found on the large plasmids. The functions of these expanded gene families have

recently been the subject of genetic investigations. Some studies have pointed to the requirement of expanded gene families and some plasmid-borne genes for response to dynamic environments inhabited by these organisms, including variations in salinity, desiccation, radiation, temperature, and other factors. The large plasmids of haloarchaea likely serve as reservoirs of paralogous genes, some of which may evolve novel functions.

Extrachromosomally encoded genes from some expanded families (*e.g.*, *orc1*/*cdc6* homologs and *tbp* homologs in *Halobacterium* sp. NRC-1) have been shown to be essential in several genetic studies, indicating that the large plasmids of *Halobacterium* sp. NRC-1 do play crucial roles in cell survival (Berquist et al., 2007; Coker and DasSarma, 2007; Facciotti et al., 2007). Clustering analysis of the large haloarchaeal protein families has allowed us to predict which members are orthologs and which may be essential, including several that are coded on plasmids (Figs. 2–4). Similar studies are not yet available for a majority of the other members of the haloarchaeal family (DasSarma and DasSarma, 2008).

The availability of multiple sequenced genomes representing ten different genera from a novel and interesting microbial family in the archaeal domain of life has provided the opportunity to understand haloarchaeal biology from a genomic perspective. A large database of information (HaloWeb database at http://halo4.umbi.umd.edu) has been used for comparative genomic analysis (DasSarma et al., Submitted for publication), which has resulted in many interesting findings. With the comparative genomic analysis of the information transfer system of haloarchaea, and extrapolation and generalization of genetic and biochemical studies from well-utilized model organisms such as *Halobacterium* sp. NRC-1 and *H. volcanii*, we now have a better understanding of the fundamental biology of the haloarchaeal family of organisms (Slonczewski et al., 2010).

Acknowledgments

The work in our laboratory is supported by the Henry M Jackson Foundation grant HU0001-09-1-0002-660883 and the National Aeronautics and Space Administration grant NNX09AC68G, with additional support provided by National Science Foundation grant MCB-0450695 (to SD). JK was supported by grant #20088033-1 from the MLTM, Republic of Korea.

References

Altman-Price, N., Mevarech, M., 2009. Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon *Haloferax volcanii*. J. Bacteriol. 191, 1610–1617.

Andersen, G.R., Valente, L., Pedersen, L., Kinzy, T.G., Nyborg, J., 2001. Crystal structures of nucleotide exchange intermediates in the eEF1A–eEF1Bα complex. Nat. Struct. Biol. 8, 531–534.

Bakke, P., Carney, N., Deloache, W., Gearing, M., Ingvorsen, K., Lotz, M., McNair, J., Penumetcha, P., Simpson, S., Voss, L., Win, M., Heyer, L.J., Campbell, A.M., 2009. Evaluation of three automated genome annotations for *Halorhabdus utahensis*. PLoS One 4, e6291.

Baliga, N.S., Bonneau, R., Facciotti, M.T., Pan, M., Glusman, G., Deutsch, E.W., Shannon, P., Chiu, Y., Weng, R.S., Gan, R.R., Hung, P., Date, S.V., Marcotte, E., Hood, L., Ng, V., 2004. Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. Genome Res. 14, 2221–2234.

Barthelmebs, L., Lecomte, B., Divies, C., Cavin, J.F., 2000. Inducible metabolism of phenolic acids in *Pediococcus pentosaceus* is encoded by an autoregulated operon which involves a new class of negative transcriptional regulator. J. Bacteriol. 182, 6724–6731.

Benelli, D., Marzi, S., Mancone, C., Alonzi, T., la Teana, A., Londei, P., 2009. Function and ribosomal localization of aIF6, a translational regulator shared by archaea and eukarya. Nucleic Acids Res. 37, 256–267.

Berquist, B., DasSarma, S., 2008. DNA replication in Archaea. In: Blum, P. (Ed.), Archaea: New Models for Prokaryotic Biology. Caister Academic Press, Norfolk, UK, pp. 95–120.

Berquist, B.R., DasSarma, P., DasSarma, S., 2007. Essential and non-essential DNA replication genes in the model halophilic Archaeon *Halobacterium* sp. NRC-1. BMC Genet. 8, 31.

Berquist, B.R., DasSarma, S., 2003. An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. J. Bacteriol. 185, 5959–5966.

Berquist, B.R., Soneja, J., DasSarma, S., 2005. Comparative genomic survey of information transfer systems in two diverse extremely halophilic archaea, *Halobacterium* sp. NRC-1 and *Haloarcula marismortui*. In: Gunde-Cimerman, N., Oren, A., Plemenitas, A. (Eds.), Adaptation to Life at High Salt Concentrations in Archaea, Bacterial, and Eukarya. Springer, Dordrecht, Netherlands, pp. 148–182.

Betlach, M.C., Shand, R.F., Leong, D.M., 1989. Regulation of the bacterioopsin gene of a halophilic archaebacterium. Can. J. Microbiol. 35, 134–140.

Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F., Oesterhelt, D., 2006. The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. BMC Genomics. 7, 169.

Boubriak, I., Ng, W.L., DasSarma, P., DasSarma, S., Crowley, D.J., McCready, S.J., 2008. Transcriptional responses to biologically relevant doses of UV-B radiation in the model archaeon, *Halobacterium* sp. NRC-1. Saline Systems 4, 13–14.

Brochier, C., Lopez-Garcia, P., Moreira, D., 2004. Horizontal gene transfer and archaeal origin of deoxyhypusine synthase homologous genes in bacteria. Gene. 330, 169–176.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S., Venter, J.C., 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273, 1058–1073.

Busch, C.R., DiRuggiero, J., 2010. MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon *Halobacterium salinarum* NRC-1. PLoS One 5, e9045.

Cheeseman, J.M., Herendeen, L.B., Cheeseman, A.T., Clough, B.F., 1997. Photosynthesis and photoprotection in mangrove under field conditions. Plant Cell Environ. 20, 579–588.

Chen, Y.H., Kocherginskaya, S.A., Lin, Y., Sriratana, B., Lagunas, A.M., Robbins, J.B., Mackie, R.I., Cann, I.K., 2005. Biochemical and mutational analyses of a unique clamp loader complex in the archaeon *Methanosarcina acetivorans*. J. Biol. Chem. 280, 41852–41863.

Coin, F., Oksenych, V., Egly, J.M., 2007. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol. Cell 26, 245–256.

Coker, J.A., DasSarma, P., Capes, M., Wallace, T., McGarrity, K., Gessler, R., Liu, J., Xiang, H., Tatusov, R., Berquist, B.R., DasSarma, S., 2009. Multiple replication origins of *Halobacterium* sp. strain NRC-1: properties of the conserved *orc7*-dependent *oriC1*. J. Bacteriol. 191, 5253–5261.

Coker, J.A., DasSarma, S., 2007. Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. BMC Genet. 8, 61.

Corbett, K.D., Berger, J.M., 2003. Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. EMBO J. 22, 151–163.

Courcelle, C.T., Chow, K.H., Casey, A., Courcelle, J., 2006. Nascent DNA processing by RecJ favors lesion repair over translesion synthesis at arrested replication forks in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 103, 9154–9159.

Crowley, D.J., Boubriak, I., Berquist, B.R., Clark, M., Richard, E., Sullivan, L., DasSarma, S., McCready, S., 2006. The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp NRC-1. Saline Systems 2, 11.

Daimon, K., Kawarabayasi, Y., Kikuchi, H., Sako, Y., Ishino, Y., 2002. Three proliferating cell nuclear antigen-like proteins found in the

hyperthermophilic archaeon *Aeropyrum pernix*: interactions with the two DNA polymerases. J. Bacteriol. 184, 687–694.
Daly, M.J., Ouyang, L., Fuchs, P., Minton, K.W., 1994. *In vivo* damage and *recA*-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*. J. Bacteriol. 176, 3508–3517.
DasSarma, P., DasSarma, S., 2008. On the origin of prokaryotic "species": the taxonomy of halophilic Archaea. Saline Systems 4, 5.
DasSarma, S., 2004. Genome sequence of an extremely halophilic archaeon. In: Fraser, C.M., Read, T.D., Nelson, K.E. (Eds.), Microbial Genomes. Humana Press, Inc., Totowa, NJ, pp. 383–399.
DasSarma, S., Capes, M., DasSarma, P., 2008. Haloarchaeal megaplasmid. In: Schwartz, E. (Ed.), Megaplasmids. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG, Berlin, pp. 3–30.
DasSarma, S., Coker, J.A., DasSarma, P., 2009. Archaea (Overview). In: Schaechter, M. (Ed.), Desk Encyclopedia of Microbiology. Academic Press, Incorporated, San Diego, pp. 118–139.
DasSarma, S., DasSarma, P., 2006. Halophiles. John Wiley & Sons Ltd.
DasSarma, S., Kennedy, S.P., Berquist, B., Victor, W., Baliga, N.S., Spudich, J.L., Krebs, M.P., Eisen, J.A., Johnson, C.H., Hood, L., 2001. Genomic perspective on the photobiology of *Halobacterium* species NRC-1, a phototrophic, phototactic, and UV-tolerant haloarchaeon. Photosynth. Res. 70, 3–17.
DasSarma, S.L., Capes, M., DasSarma, P., DasSarma, S., 2010. HaloWeb: the haloarchaeal genomes database. Saline Systems 6, 12.
Denver, D.R., Swenson, S.L., Lynch, M., 2003. An evolutionary analysis of the helix–hairpin-helix superfamily of DNA repair glycosylases. Mol. Biol. Evol. 20, 1603–1611.
DeVeaux, L.C., Müller, J.A., Smith, J., Petrisko, J., Wells, D.P., DasSarma, S., 2007. Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression. Radiat. Res. 168, 507–514.
Diaz-Lazcoz, Y., Aude, J.C., Nitschke, P., Chiapello, H., Landes-Devauchelle, C., Risler, J.L., 1998. Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases. Mol. Biol. Evol. 15, 1548–1561.
DiRuggiero, J., Santangelo, N., Nackerdien, Z., Ravel, J., Robb, F.T., 1997. Repair of extensive ionizing-radiation DNA damage at 95 degrees C in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. 179, 4643–4645.
Duggin, I.G., McCallum, S.A., Bell, S.D., 2008. Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. Proc. Natl. Acad. Sci. USA 105, 16737–16742.
Facciotti, M.T., Reiss, D.J., Pan, M., Kaur, A., Vuthoori, M., Bonneau, R., Shannon, P., Srivastava, A., Donohoe, S.M., Hood, L.E., Baliga, N.S., 2007. General transcription factor specified global gene regulation in archaea. Proc. Natl. Acad. Sci. USA 104, 4630–4635.
Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J., Oesterhelt, D., 2005. Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. Genome Res. 15, 1336–1343.
Farhoud, M.H., Wessels, H.J., Steenbakkers, P.J., Mattijssen, S., Wevers, R.A., van Engelen, B.G., Jetten, M.S., Smeitink, J.A., van den Heuvel, L.P., Keltjens, J.T., 2005. Protein complexes in the archaeon *Methanothermobacter thermautotrophicus* analyzed by blue native/SDS–PAGE and mass spectrometry. Mol. Cell Proteomics 4, 1653–1663.
Forterre, P., Gadelle, D., 2009. Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. Nucleic Acids Res. 37, 679–692.
Forterre, P., Gribaldo, S., Gadelle, D., Serre, M.C., 2007. Origin and evolution of DNA topoisomerases. Biochimie. 89, 427–446.
Franzmann, P.D., Stackebrandt, E., Sanderson, K., Volkman, J.K., Cameron, D.E., Stevenson, P.L., McMeekin, T.A., Burton, H.R., 1988. *Halobacterium lacusprofundi* sp. nov., a halophilic bacterium isolated from Deep Lake, Antarctica. Syst. Appl. Microbiol. 11, 20–27.
Gaudier, M., Schuwirth, B.S., Westcott, S.L., Wigley, D.B., 2007. Structural basis of DNA replication origin recognition by an ORC protein. Science 317, 1213–1216.
Grunberg, S., Bartlett, M.S., Naji, S., Thomm, M., 2007. Transcription factor E is a part of transcription elongation complexes. J. Biol. Chem. 282, 35482–35490.
Hartman, A.L., Norais, C., Badger, J.H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T.M., Maupin-Furlow, J., Pohlschroder, M., Daniels, C., Pfeiffer, F., Allers, T., Eisen, J.A., 2010. The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. PLoS One 5, e9605.
Hasenöhrl, D., Benelli, D., Barbazza, A., Londei, P., Blasi, U., 2006. *Sulfolobus solfataricus* translation initiation factor 1 stimulates translation initiation complex formation. RNA 12, 674–682.
Hasenöhrl, D., Fabbretti, A., Londei, P., Gualerzi, C.O., Blasi, U., 2009. Translation initiation complex formation in the crenarchaeon *Sulfolobus solfataricus*. RNA 15, 2288–2298.
Heinze, R.J., Giron-Monzon, L., Solovyova, A., Elliot, S.L., Geisler, S., Cupples, C.G., Connolly, B.A., Friedhoff, P., 2009. Physical and functional interactions between *Escherichia coli* MutL and the Vsr repair endonuclease. Nucleic Acids Res. 37, 4453–4463.
Hendrickson, E.L., Kaul, R., Zhou, Y., Bovee, D., Chapman, P., Chung, J., Conway de Macario, E., Dodsworth, J.A., Gillett, W., Graham, D.E., Hackett, M., Haydock, A.K., Kang, A., Land, M.L., Levy, R., Lie, T.J., Major, T.A., Moore, B.C., Porat, I., Palmeiri, A., Rouse, G., Saenphimmachak, C., Soll, D., Van Dien, S., Wang, T., Whitman, W.B., Xia, Q., Zhang, Y., Larimer, F.W., Olson, M.V., Leigh, J.A., 2004. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. J. Bacteriol. 186, 6956–6969.
Henneke, G., Flament, D., Hubscher, U., Querellou, J., Raffin, J.P., 2005. The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. J. Mol. Biol. 350, 53–64.
Henneke, G., Gueguen, Y., Flament, D., Azam, P., Querellou, J., Dietrich, J., Hubscher, U., Raffin, J.P., 2002. Replication factor C from the hyperthermophilic archaeon *Pyrococcus abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RFC PCNA-binding domain. J. Mol. Biol. 323, 795–810.
Hescox, M.A., Carlberg, D.M., 1972. Photoreactivation in *Halobacterium cutirubrum*. Can. J. Microbiol. 18, 981–985.
Holliday, R., 1964. A mechanism for gene conversion in fungi. Genet. Res. Camb. 5, 282–304.
Hopfner, K.P., Karcher, A., Shin, D., Fairley, C., Tainer, J.A., Carney, J.P., 2000. Mre11 and Rad50 from *Pyrococcus furiosus*: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine. J. Bacteriol. 182, 6036–6041.
Hyland, E.M., Cosgrove, M.S., Molina, H., Wang, D., Pandey, A., Cottee, R.J., Boeke, J.D., 2005. Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. Mol. Cell Biol. 25, 10060–10070.
Inagaki, Y., Doolittle, F.W., 2000. Evolution of the eukaryotic translation termination system: origins of release factors. Mol. Biol. Evol. 17, 882–889.
Kanai, S., Kikuno, R., Toh, H., Ryo, H., Todo, T., 1997. Molecular evolution of the photolyase-blue-light photoreceptor family. J. Mol. Evol. 45, 535–548.
Kanugula, S., Pauly, G.T., Moschel, R.C., Pegg, A.E., 2005. A bifunctional DNA repair protein from *Ferroplasma acidarmanus* exhibits O6-alkylguanine-DNA alkyltransferase and endonuclease V activities. Proc. Natl. Acad. Sci. USA 102, 3617–3622.
Kasiviswanathan, R., Shin, J.H., Kelman, Z., 2006. DNA binding by the *Methanothermobacter thermautotrophicus* Cdc6 protein is inhibited by the minichromosome maintenance helicase. J. Bacteriol. 188, 4577–4580.
Kelman, Z., Hurwitz, J., 2000. A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* ΔH. J. Biol. Chem. 275, 7327–7336.
Kennedy, S., (2003). Understanding genome structure, function, and evolution in the halophilic archaeon *Halobacterium* NRC-1. Department of Microbiology. Amherst, University of Massachusetts: 171.
Kish, A., DiRuggiero, J., 2008. Rad50 is not essential for the Mre11-dependent repair of DNA double-strand breaks in *Halobacterium* sp. strain NRC-1. J. Bacteriol. 190, 5210–5216.
Kiyonari, S., Tahara, S., Shirai, T., Iwai, S., Ishino, S., Ishino, Y., 2009a. Biochemical properties and base excision repair complex formation of apurinic/apyrimidinic endonuclease from *Pyrococcus furiosus*. Nucleic Acids Res. 37, 6439–6453.
Kiyonari, S., Tahara, S., Uchimura, M., Shirai, T., Ishino, S., Ishino, Y., 2009b. Studies on the base excision repair (BER) complex in *Pyrococcus furiosus*. Biochem. Soc. Trans. 37, 79–82.
Kunkel, T.A., Erie, D.A., 2005. DNA mismatch repair. Annu. Rev. Biochem. 74, 681–710.
Kuzminov, A., 2001. DNA replication meets genetic exchange: chromosomal damage and its repair by homologous recombination. Proc. Natl. Acad. Sci. USA 98, 8461–8468.
Lam, W.L., Doolittle, W.F., 1989. Shuttle vectors for the archaebacterium *Halobacterium volcanii*. Proc. Natl. Acad. Sci. USA 86, 5478–5482.
Lange, U., Hausner, W., 2004. Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. Mol. Microbiol. 52, 1133–1143.
Lao-Sirieix, S.H., Bell, S.D., 2004. The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA

and RNA primase, polymerase and 3′-terminal nucleotidyl transferase activities. J. Mol. Biol. 344, 1251–1263.
Lao-Sirieix, S.-H., Marsh, V.L., Bell, S.D., 2007. DNA replication and cell cycle. In: Cavicchioli, R. (Ed.), Archaea: Molecular and Cellular Biology. ASM Press, Washington, DC, pp. 93–109.
Le Breton, M., Henneke, G., Norais, C., Flament, D., Myllykallio, H., Querellou, J., Raffin, J.P., 2007. The heterodimeric primase from the euryarchaeon *Pyrococcus abyssi*: a multifunctional enzyme for initiation and repair? J. Mol. Biol. 374, 1172–1185.
Lindas, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J., Bernander, R., 2008. A unique cell division machinery in the Archaea. Proc. Natl. Acad. Sci. USA 105, 18942–18946.
Liu, J., Lou, Y., Yokota, H., Adams, P.D., Kim, R., Kim, S.H., 2005. Crystal structure of a PhoU protein homologue: a new class of metalloprotein containing multinuclear iron clusters. J. Biol. Chem. 280, 15960–15966.
Liu, L., Komori, K., Ishino, S., Bocquier, A.A., Cann, I.K., Kohda, D., Ishino, Y., 2001. The archaeal DNA primase: biochemical characterization of the p41–p46 complex from *Pyrococcus furiosus*. J. Biol. Chem. 276, 45484–45490.
Liu, Y., West, S.C., 2004. Happy Hollidays: 40th anniversary of the Holliday junction. Nat. Rev. Mol. Cell Biol. 5, 937–944.
Luijsterburg, M.S., White, M.F., van Driel, R., Dame, R.T., 2008. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. Crit. Rev. Biochem. Mol. Biol. 43, 393–418.
Lundgren, M., Bernander, R., 2007. Genome-wide transcription map of an archaeal cell cycle. Proc. Natl. Acad. Sci. USA 104, 2939–2944.
Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B., Koonin, E.V., 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res. 30, 482–496.
Makarova, K.S., Sorokin, A.V., Novichkov, P.S., Wolf, Y.I., Koonin, E.V., 2007. Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. Biol. Direct. 2, 33.
Maki, H., Sekiguchi, M., 1992. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. Nature 355, 273–275.
Malfatti, S., Tindall, B., Schneider, S., Fähnrich, R., Lapidus, A., Labutti, K., Copeland, A., Glavina del Rio, T., Nolan, M., Chen, F., Lucas, S., Tice, H., Cheng, J.-F., Bruce, D., Goodwin, L., Pitluck, S., Anderson, I., Pati, A., Ivanova, N., Mavrommatis, K., Chen, A., Palaniappan, K., D'Haeseleer, P., Göker, M., Bristow, J., Eisen, J., Markowitz, V., Hugenholtz, P., Kyrpides, N., Klenk, H., Chain, P., 2009. Complete genome sequence of *Halogeometricum borinquense* type strain ($PR3^T$). Stand. Genomic Sci. 1.
Manzur, K.L., Zhou, M.M., 2005. An archaeal SET domain protein exhibits distinct lysine methyltransferase activity towards DNA-associated protein MC1-α. FEBS Lett. 579, 3859–3865.
Marintchev, A., Wagner, G., 2004. Translation initiation: structures, mechanisms and evolution. Q Rev. Biophys. 37, 197–284.
Matsunaga, F., Forterre, P., Ishino, Y., Myllykallio, H., 2001. In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc. Natl. Acad. Sci. USA 98, 11152–11157.
May, B.P., Dennis, P.P., 1989. Evolution and regulation of the gene encoding superoxide dismutase from the archaebacterium *Halobacterium cutirubrum*. J. Biol. Chem. 264, 12253–12258.
McCready, S., Marcello, L., 2003. Repair of UV damage in *Halobacterium salinarum*. Biochem. Soc. Trans. 31, 694–698.
McCready, S., Müller, J.A., Boubriak, I., Berquist, B.R., Ng, W.L., DasSarma, S., 2005. UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium* sp NRC-1. Saline Systems 1, 3.
Mitchell, S.F., Lorsch, J.R., 2008. Should I stay or should I go? Eukaryotic translation initiation factors 1 and 1A control start codon recognition. J. Biol. Chem. 283, 27345–27349.
Moser, J., Lorenz, S., Hubschwerlen, C., Rompf, A., Jahn, D., 1999. *Methanopyrus kandleri* glutamyl-tRNA reductase. J. Biol. Chem. 274, 30679–30685.
Müller, J.A., DasSarma, S., 2005. 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, 1659–1667.
Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., Forterre, P., 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science 288, 2212–2215.
Ng, W.L., DasSarma, S., 1993. Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. J. Bacteriol. 175, 4584–4596.
Ng, W.V., Berquist, B.R., Coker, J.A., Capes, M., Wu, T.H., Dassarma, P., Dassarma, S., 2008. Genome sequences of *Halobacterium* species. Genomics 91, 548–552. author reply 553-4.
Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroder, M., Spudich, J.L., Jung, K.W., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L., DasSarma, S., 2000. Genome sequence of *Halobacterium* species NRC-1. Proc. Natl. Acad. Sci. USA 97, 12176–12181.
Nichols, M.D., DeAngelis, K., Keck, J.L., Berger, J.M., 1999. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. EMBO J. 18, 6177–6188.
Nishino, T., Komori, K., Ishino, Y., Morikawa, K., 2001. Dissection of the regional roles of the archaeal Holliday junction resolvase Hjc by structural and mutational analyses. J. Biol. Chem. 276, 35735–35740.
Nomura, M., Gourse, R., Baughman, G., 1984. Regulation of the synthesis of ribosomes and ribosomal components. Annu. Rev. Biochem. 53, 75–117.
Norais, C., Hawkins, M., Hartman, A.L., Eisen, J.A., Myllykallio, H., Allers, T., 2007. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. PLoS Genet. 3, e77.
O'Connell, A., Robin, G., Kobe, B., Botella, J.R., 2009. Biochemical characterization of *Arabidopsis* developmentally regulated G-proteins (DRGs). Protein Expr. Purif. 67, 88–95.
Ohtani, N., Yanagawa, H., Tomita, M., Itaya, M., 2004. Identification of the first archaeal Type 1 RNase H gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA–DNA junction. Biochem. J. 381, 795–802.
Patel, S.S., Picha, K.M., 2000. Structure and function of hexameric helicases. Annu. Rev. Biochem. 69, 651–697.
Pestova, T.V., Borukhov, S.I., Hellen, C.U., 1998. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. Nature 394, 854–859.
Pfeiffer, F., Schuster, S.C., Broicher, A., Falb, M., Palm, P., Rodewald, K., Ruepp, A., Soppa, J., Tittor, J., Oesterhelt, D., 2008. Evolution in the laboratory: the genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. Genomics 91, 335–346.
Pisani, F.M., De Felice, M., Carpentieri, F., Rossi, M., 2000. Biochemical characterization of a clamp-loader complex homologous to eukaryotic replication factor C from the hyperthermophilic archaeon *Sulfolobus solfataricus*. J. Mol. Biol. 301, 61–73.
Ren, B., Kuhn, J., Meslet-Cladiere, L., Myllykallio, H., Ladenstein, R., 2007. Crystallization and preliminary X-ray analysis of a RecB-family nuclease from the archaeon *Pyrococcus abyssi*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 63, 406–408.
Robey-Bond, S.M., Barrantes-Reynolds, R., Bond, J.P., Wallace, S.S., Bandaru, V., 2008. *Clostridium acetobutylicum* 8-oxoguanine DNA glycosylase (Ogg) differs from eukaryotic Oggs with respect to opposite base discrimination. Biochemistry 47, 7626–7636.
Robinson, N.P., Bell, S.D., 2007. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. Proc. Natl. Acad. Sci. USA 104, 5806–5811.
Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R., Bell, S.D., 2004. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell 116, 25–38.
Rouillon, C., White, M.F., 2010. The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair. J. Biol. Chem. 285, 11013–11022.
Rudolf, J., Rouillon, C., Schwarz-Linek, U., White, M.F., 2010. The helicase XPD unwinds bubble structures and is not stalled by DNA lesions removed by the nucleotide excision repair pathway. Nucleic Acids Res. 38, 931–941.
Samson, R., Reeve, J.N., 2007. DNA binding proteins and chromatin. In: Cavicchioli, R. (Ed.), Archaea: Molecular and Cellular Biology. ASM Press, Washington, DC, pp. 110–119.
Sartori, A.A., Fitz-Gibbon, S., Yang, H., Miller, J.H., Jiricny, J., 2002. A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. EMBO J. 21, 3182–3191.
Saunders, E., Tindall, B.J., Fähnrich, R., Lapidus, A., Copeland, A., Rio, T.G.D., Lucas, S., Chen, F., Tice, H., Cheng, J.-F., Han, C., Detter, J.C., Bruce, D., Goodwin, L., Chain, P., Pitluck, S., Pati, A., Ivanova, N., Mavromatis, K., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Brettin, T., Rohde, M., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.-P., Kyrpides, N.C., 2010. Complete genome sequence of *Haloterrigena turkmenica* type strain ($4k^T$). Stand. Genomic Sci. 2, 107–116.
Seybert, A., Scott, D.J., Scaife, S., Singleton, M.R., Wigley, D.B., 2002. Biochemical characterisation of the clamp/clamp loader proteins from the euryarchaeon *Archaeoglobus fulgidus*. Nucleic Acids Res. 30, 4329–4338.

Sheppard, K., Söll, D., 2008. On the evolution of the tRNA-dependent amidotransferases, GatCAB and GatDE. J. Mol. Biol. 377, 831–844.
Shioda, M., Sugimori, K., Shiroya, T., Takayanagi, S., 1989. Nucleosomelike structures associated with chromosomes of the archaebacterium *Halobacterium salinarium*. J. Bacteriol. 171, 4514–4517.
Si, K., Maitra, U., 1999. The *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 6 does not function as a translation initiation factor. Mol. Cell Biol. 19, 1416–1426.
Slonczewski, J.L., Coker, J.A., DasSarma, S., 2010. Microbial Growth with Multiple Stressors. Microbe 5, 110–115.
Soliman, G.S.H., Truper, H.G., 1982. *Halobacterium pharaonis* sp. nov., a new, extremely haloalkaliphilic archaebacterium with low magnesium requirement. Zbl. Bakt. Hyg. I Abt. Orig. 3, 318–329.
Takayanagi, S., Morimura, S., Kusaoke, H., Yokoyama, Y., Kano, K., Shioda, M., 1992. Chromosomal structure of the halophilic archaebacterium *Halobacterium salinarium*. J. Bacteriol. 174, 7207–7216.
Tindall, B.J., Schneider, S., Lapidus, A., Copeland, A., Rio, T.G.D., Nolan, M., Lucas, S., Chen, F., Tice, H., Cheng, J.-F., Saunders, E., Bruce, D., Goodwin, L., Pitluck, S., Mikhailova, N., Pati, A., Ivanova, N., Mavrommatis, K., Chen, A., Palaniappan, K., Chain, P., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Brettin, T., Han, C., Rohde, M., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.-P., Kyrpides, N.C., Detter, J.C., 2009. Complete genome sequence of *Halomicrobium mukohataei* type strain (arg-2^T). Stand. Genomic Sci. 1.
Walter, P., Klein, F., Lorentzen, E., Ilchmann, A., Klug, G., Evguenieva-Hackenberg, E., 2006. Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol. Microbiol. 62, 1076–1089.
Wang, J., Dasgupta, I., Fox, G.E., 2009. Many nonuniversal archaeal ribosomal proteins are found in conserved gene clusters. Archaea 2, 241–251.
White, M.F., Bell, S.D., 2002. Holding it together: chromatin in the Archaea. Trends Genet. 18, 621–626.
Williams, G.J., Johnson, K., Rudolf, J., McMahon, S.A., Carter, L., Oke, M., Liu, H., Taylor, G.L., White, M.F., Naismith, J.H., 2006. Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 62, 944–948.
Winter, J.A., Christofi, P., Morroll, S., Bunting, K.A., 2009. The crystal structure of *Haloferax volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. BMC Struct. Biol. 9, 55.
Woods, W.G., Dyall-Smith, M.L., 1997. Construction and analysis of a recombination-deficient (*radA*) mutant of *Haloferax volcanii*. Mol. Microbiol. 23, 791–797.
Wu, J., Rosen, B.P., 1993. Metalloregulated expression of the *ars* operon. J. Biol. Chem. 268, 52–58.
Zhao, A., Gray, F.C., MacNeill, S.A., 2006. ATP- and NAD$^+$-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. Mol. Microbiol. 59, 743–752.
Zou, Y., Liu, Y., Wu, X., Shell, S.M., 2006. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. J. Cell Physiol. 208, 267–273.