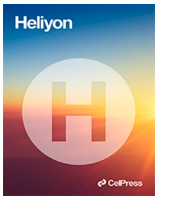
Heliyon 10 (2024) e27382

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Review article

Restriction modification systems in archaea: A panoramic outlook

Pallavi Gulati [a,*], Ashish Singh [b], Sandeep Patra [a,1], Shreyas Bhat [a,1], Anil Verma [c]

[a] Department of Microbiology, Ram Lal Anand College, University of Delhi South Campus, New Delhi 110021, India
[b] Department of Microbiology, University of Delhi South Campus, New Delhi 110021, India
[c] Department of Pathology, School of Medicine, University of Pittsburgh, Pittsburgh, PA-15213, USA

ARTICLE INFO

Keywords:
Restriction-modification system
Archaea
DNA methylation
Defense response
Endonuclease
Methyltransferase

ABSTRACT

Restriction modification (RM) systems are one of the ubiquitous yet primitive defense responses employed by bacteria and archaea with the primary role of safeguarding themselves against invading bacteriophages. Protection of the host occurs by the cleavage of the invading foreign DNA via restriction endonucleases with concomitant methylation of host DNA with the aid of a methyltransferase counterpart. RM systems have been extensively studied in bacteria, however, in the case of archaea there are limited reports of RM enzymes that are investigated to date owing to their inhospitable growth demands. This review aims to broaden the knowledge about what is known about the diversity of RM systems in archaea and encapsulate the current knowledge on restriction and modification enzymes characterized in archaea so far and the role of RM systems in the milieu of archaeal biology.

1. Introduction

The restriction-modification (RM) system is a primary innate immune response found in bacteria and archaea. The phenomenon of a systems modification system is laid by an interesting concept of host-controlled variation in bacterial viruses. Bertani and Weigle demonstrated this phenomenon in a one-step growth experiment by cycling phage P2 in two strains- *E. coli* and *Shigella*, as well as cycling phage λ in *E. coli* K12 and *E. coli* C [1]. Host-controlled variation refers to the phenomenon wherein one cycle of growth of a particular phage on a certain host alters the capacity of progeny phage to propagate on other hosts. Luria and Human also documented the presence of this infection barrier when B/4 mutants of *E. coli* were exposed to T-even phages [2]. Consequently, the phenomenon of host-controlled variation is recognized as a distinctive defense mechanism employed by bacteria to counteract bacteriophage attacks. RM systems protect the host DNA by restricting the invasion of foreign DNA and simultaneously methylating the host DNA, which makes the latter resistant to cleavage.

RM systems can be classified into four types based on their subunit composition, sequence recognition, cleavage pattern, and cofactor requirements [3–5]. Type I RM systems are multi-subunit enzymes and consist of three subunits: hsdR, hsdS, and hsdM. These subunits pair up to perform two antagonistic functions, depending on their oligomeric state and the methylation status of the target DNA. The pentamer form of R2M2S acts as a restriction endonuclease (REase), which cleaves unmethylated or inappropriately methylated DNA from bacteriophages, using ATP, Mg^{2+}, and AdoMet as cofactors. In contrast, the heterotrimer form of M2S acts as a methyltransferase (MTase), that utilizes AdoMet as a methyl donor that transfers the methyl group to the target base, thereby

* Corresponding author.
E-mail address: pallavi6iitd@gmail.com (P. Gulati).
[1] These two authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e27382
Received 22 August 2023; Received in revised form 19 February 2024; Accepted 28 February 2024
Available online 5 March 2024

protecting the host DNA from cleavage. The recognition sequence of Type I RM systems is asymmetric and bipartite, consisting of a 3–4 base pair sequence at the 5′ end and a 4–5 base pair sequence at the 3′ end, separated by a non-specific spacer sequence (e.g., EcoKI: 5′-AACN6GTGC). Type I RM systems cleave DNA approximately 1000 base pairs away from the recognition sequence [6,7]. On the other hand, Type II RM systems have separate endonuclease and methyltransferase components that catalyze restriction and methylation independently. The recognition sequence for Type II RM systems is usually palindromic and ranges from 4 to 8 base pairs. The endonuclease component cleaves the DNA within this recognition sequence, except for Type IIS systems that have a different cleavage pattern (e.g., EcoKI: 5′-GAATTC) [8–10]. There are distinct subtypes within Type II enzymes that exhibits unique cleavage characteristics. Type IIG enzymes recognizes sequences in either a symmetric or asymmetric fashion. Type IIM enzymes recognize methylated sequences, while Type IIL enzymes cleave DNA specifically 20 bases away from the recognition site. Type II enzymes are the most diverse group of restriction enzymes and are known as workhorses of molecular biology with numerous applications in cloning, DNA fingerprinting, gene mapping, etc. Type III RM systems consist of Mod and Res subunits, that form a functional holo-enzyme. The (Mod)2 subunit is responsible for methylation, while the restriction is catalyzed by Res2Mod2. These systems recognize asymmetric bipartite sequences and cleave the DNA 25–27 base pairs downstream of the recognition site [11–14]. The defense system of RM is supported by the fact that many bacteriophages develop strategies to counter the fortification offered by the RM system. To protect the bacteria from such bacteriophages with the modified genome, special types of RM systems such as Type IV and Type IIM) have evolved that can hydrolyze only methylated DNA [7,8,10,15–17]. Fig. 1 illustrates the mechanism of action of the four types of RM enzymes.

Since Type I restriction enzymes cleave 1000 base pairs away from the recognition sequence, identifying their recognition sites has been challenging. The recognition site of EcoKI was determined by mapping wild type strains as well as mutant strains of Φx172, generated using marker rescue technique that confers sensitivity to EcoKI. The consensus sequence obtained from the EcoKI-sensitive strains was also searched in pBR322, G4, SV40 and bacteriophage fd, using a computer-based search program that was utilized to examine the region of homology in all strains sensitive to EcoKI restriction [18]. Similarly, the recognition site of EcoB was examined by comparing the genome sequences of ΦXsB1, SV40 and f1 bacteriophage via computer-based search and mutational studies [19]. The recognition sequence of EcoD methyltransferase was elucidated by using DNA of known sequences that were methylated by EcoD methyltransferase using 3H AdoMet as methyl donor. The methylated sites within the recognition sequence were mapped by digestion of the methylated fragment with Type II restriction enzymes. This was followed by gel electrophoresis and fluorography to detect the methylation in DNA. The mapped sites were further fed to the computer program to search for the consensus sequence present in all labeled regions [20]. The emergence of novel sequencing technologies, such as single-molecule real-time (SMRT) sequencing of Pacific Biosciences, has facilitated the characterization of methylated motifs during genome mapping at a large scale in a convenient manner. SMRT sequencing uses fluorescently tagged nucleotides for detecting methylated bases [21]. It involves a SMRT cell with tiny pores called ZMWs that immobilize DNA templates. Each nucleotide is labeled with a distinct fluorescent tag, producing a unique signal upon incorporation. The fluorescence pulse duration is longer for methylated adenine (3–5 times) and methylated cytosine (5–7 times) compared to their unmethylated counterparts. The RS Modification and Motif analysis program analyzes the IPD ratio, which represents the time taken to incorporate a base at a specific position compared to an unmodified base [22]. The Oxford Nanopore

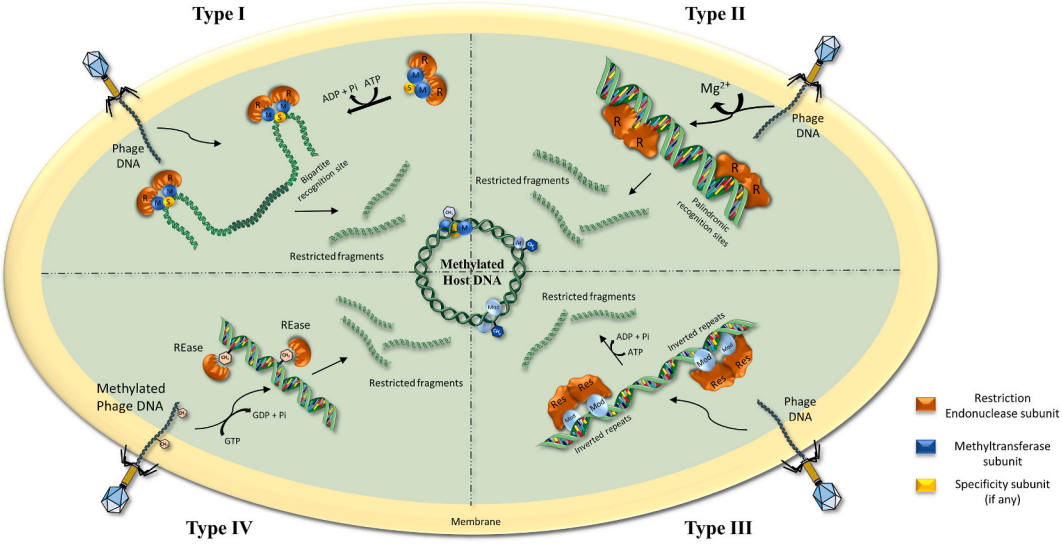


Fig. 1. Types of Restriction modification systems. Type I RM systems cleaves the phage DNA via R2M2S pentamer. M2S trimer methylates the host DNA and renders it immune against self-cleavage. Type II restriction enzymes cleaves the phage DNA via R2 homodimer and methylation occurs via monomeric M subunit. Restriction by Type III enzymes occurs via Res2Mod2 and self DNA is methylated by Mod2 homodimer. Type IV enzymes cleaves the methylated DNA of incoming phage particles. Methylation of host DNA by methyltransferase prohibits the action of cognate restriction endonuclease and thus protects the host DNA from cleavage.

Table 1
List of all the archaeal species with known recognition sequence and methylated base (http://rebase.neb.com).

Organism Name	Restriction Enzyme	Type	Predicted Recognition Sequence	Predicted Methylation Product	Phylum	Reference
Aeropyrum pernix K1	M.*Ape*KI	Type II	GCWGC	m5C	Thermoproteota	(Hayashi et al., 2021; Kawarabayasi, 1999)
Cenarchaeum symbiosum A	M.CsyAORF1631P	Type II	CCWGG	m5C	Thaumarchaeota	(Hallam et al., 2006)
Haloarcula sinaiiensis	Hsi33800V	Type II	CTCCAG	m6A	Euryarchaeota	http://rebase.neb.com
Haloarcula sinaiiensis	M.Hsi33800I	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com
Haloarcula sinaiiensis	M.Hsi33800II	Type II	GATC	m6A	Euryarchaeota	http://rebase.neb.com
Halobacterium salinarum R1	S.HsaR1ORF1171P	Type I	CAGNNNNNNTGCT	–	Euryarchaeota	(Pfeiffer et al., 2008)
Halobacterium species BOL4-2	M.Hsp42I	Type I	CAGNNNNNNTGCT	m6A	Euryarchaeota	http://rebase.neb.com
Halobacterium species GSL-19	HspGSL19II	Type II	GTCCAG	m6A	Euryarchaeota	(DasSarma et al., 2019)
Halobacterium species GSL-19	M.HspGSL19I	Type II	CTAG	m4C	Euryarchaeota	(DasSarma et al., 2021)
Halobacterium species NRC-1	M.HspNII	Type II	CTAG	m4C	Euryarchaeota	(DasSarma et al., 2022)
Halobacterium species NRC-34001	M.Hsp34001I	Type II	CTAG	m4C	Euryarchaeota	(DasSarma et al., 2022)
Halobacterium species NRC-34001	M.Hsp34001ORF480P	Type I	CAGNNNNNNTGCT	m6A	Euryarchaeota	(DasSarma et al., 2022)
Haloferax mediterrreanei	M.Hme33500I	Type II	CTAG	m4C	Euryarchaeota	(DasSarma et al., 2019)
	M.Hme33500II	Type II	HGCWGCK	m4C	Euryarchaeota	(DasSarma et al., 2019)
Haloferax volcanii DS2	M.HvoDSI	Type II	CTAG	m4C	Euryarchaeota	(Harris & Goldman, 2020; Hartman et al., 2010)
Haloferax volcanii DS2	M.HvoDSII	Type I	GCABNNNNNNVTGC	m6A	Euryarchaeota	(Hartman et al., 2010; Ouellette et al., 2015)
Haloferax volcanii DS2	M.HvoDSORF40P	Type II	GTCGAC	m5C	Euryarchaeota	(Hartman et al., 2010; Ouellette et al., 2015)
Halogeometricum borinquense DSM 11551	M.Hbo11551I	Type II	CTAG	m4C	Euryarchaeota"	(Malfatti et al., 2009)
Halogranum salarium B-1	M.Hsa1ORF39130P	Type II	GTCGAC	m5C	Euryarchaeota	(Kim et al., 2005)
Halopenitus persicus CBA1233	M.HpeORFCP	Type II	GCATGC	m5C	Euryarchaeota	http://rebase.neb.com
Haloprofundus species MHR1	RM.HspMHR1II	Type II	GAGCAGC	m6A	Euryarchaeota	http://rebase.neb.com
	M.HspMHR1I	Type II	TCGWCGA	m4C	Euryarchaeota	http://rebase.neb.com
Haloquadratum walsbyi	M.*Hwa*ORF1970P	Type II	GGCC	m4C	Euryarchaeota	(Bolhuis et al., 2006)
Haloquadratum walsbyi C23	M.HwaC23ORF2138P	Type II	GGCC	m4C	Euryarchaeota	(Dyall-Smith et al., 2011)
Halorhabdus tiamatea SARL4B	M.Hti4BORF11845P	Type II	GTCGAC	m5C	Euryarchaeota	(Antunes et al., 2011)
Halorhabdus tiamatea SARL4B	M.Hti4BORF13250P	Type II	GTCGAC	m5C	Euryarchaeota	
Halorhabdus tiamatea SARL4B	M.Hti4BORF13590P	Type II	CCCGT	m5C	Euryarchaeota	
Halorhabdus tiamatea SARL4B	M.Hti4BORF752P	Type II	GTCGAC	m5C	Euryarchaeota	
Halorhabdus tiamatea SARL4B	M.Hti4BORF8385P	Type II	GTCGAC	m5C	Euryarchaeota	
Halorubrum lacusprofundi	M.Hla49239I	Type II	CCWGG	m4C	Euryarchaeota	(Anderson et al., 2016)
Halorubrum species Bo13-1	M.HspBol31I	Type I	GAANNNNNCTCC	m6A	Euryarchaeota	(Anton et al., 2021a)

(*continued on next page*)

Table 1 (*continued*)

Organism Name	Restriction Enzyme	Type	Predicted Recognition Sequence	Predicted Methylation Product	Phylum	Reference
Halorubrum species Bo13-1	M.HspBol31II	Type II	GATC	m6A	Euryarchaeota	(DasSarma et al., 2019)
Halorubrum species Bo13-1	M.HspBol31III	Type II	CTCGAG	m5C	Euryarchaeota	(Anton et al., 2021b)
Halorubrum species PV6	M.HspPV6I	Type II	GATC	m6A	Euryarchaeota	(Zaretsky et al., 2018)
Haloterrigena jeotgali A29	M.HjeA29ORF15385P	Type II	GWGCWC	m5C	Euryarchaeota	(Cha et al., 2015)
Haloterrigena salifodinae BOL5-1	M.Hsa51I	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com
Haloterrigena salifodinae BOL5-2	M.Hsa51II	Type II	CATTC	m6A	Euryarchaeota	http://rebase.neb.com
Haloterrigena turkmenica DSM 5511	M.HtuI	Type II	CTAG	m4C	Euryarchaeota	(Blow et al., 2016); (Saunders et al., 2010)
Haloterrigena turkmenica DSM 5511	M.HtuII	Type II	CATTC	m6A	Euryarchaeota	
Halovivax ruber XH-70	M.HruXH70ORF3021P	Type II	GTCGAC	m5C	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter arboriphilus SA	M.MmiSM9ORF453P	Type II	GGNCC	m5C	Euryarchaeota	(Kelly et al., 2016)
Methanobrevibacter ruminantium M1	M2.MruORF26P	Type II	CCCGC	m5C	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter ruminantium M2	M.MsmAORF531P	Type II	CCTTC	m5C	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter species YE315	M.Msp315II	Type III	CAGAAA	m6A	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter species YE316	M.Msp315ORF4580P	Type II	GGNCC	m5C	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter species YE317	M.Msp315ORF705P	Type II	GATC	m5C	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter species YE318	M1.Msp315I	Type II	GGATG	m6A	Euryarchaeota	http://rebase.neb.com
Methanobrevibacter species YE319	M2.Msp315I	Type II	GGATG	m6A	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2661	M.*Mja*I	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2662	M.*Mja*II	Type II	GGNCC	m5C	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2663	M.*Mja*III	Type II	GATC	m6A	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2664	M.*Mja*IV	Type II	GTNNAC	m6A	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2665	M.MjaIX	Type II	CCANNNNNGTR	m6A	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2666	M.*Mja*V	Type II	GTAC	m4C	Euryarchaeota	http://rebase.neb.com
Methanocaldococcus jannaschii DSM 2667	M.MjaVI	Type II	CCGG	m4C	Euryarchaeota	(Bult et al., 1996)
Methanocaldococcus jannaschii DSM 2668	M.MjaVII	Type I	CAANNNNNNNTGG	m6A	Euryarchaeota	(Blow et al., 2016)
Methanocaldococcus jannaschii DSM 2669	M.MjaVIII	Type I	GAYNNNNNGTAA	m6A	Euryarchaeota	(Blow et al., 2016)
Methanocella arvoryzae MRE50	M.Mar50ORF2617P	Type II	CGCG	m4C	Euryarchaeota	(Erkel et al., 2006)
Methanocella conradii HZ254	M.Mco254ORF415P	Type II	AGCT	m5C	Euryarchaeota	http://rebase.neb.com
Methanocella paludicola SANAE	M.MpaSORF1626P	Type II	CGCG	m4C	Euryarchaeota	http://rebase.neb.com
Methanocella paludicola SANAE	M.MpaSORF2216P	Type II	CGCG	m4C	Euryarchaeota	http://rebase.neb.com
Methanococcus maripaludis XI X1	M.MvaSBORF154P	Type III	CGCG	m6A	Euryarchaeota	http://rebase.neb.com
Methanococcus voltae A3	M.MlaZI	Type I	ACCNNNNNNRTGA	m6A	Euryarchaeota	http://rebase.neb.com
Methanococcus voltae A3	M.MlaZII	Type II	GTAC	m4C	Euryarchaeota	http://rebase.neb.com
Methanoculleus marisnigri JR1	M.MseS3FaORF11145P	Type II	CCWGG	m5C	Euryarchaeota	(S.-C. Chen et al., 2016)

(*continued on next page*)

Table 1 (*continued*)

Organism Name	Restriction Enzyme	Type	Predicted Recognition Sequence	Predicted Methylation Product	Phylum	Reference
Methanoculleus marisnigri JR1	M.MseS3FaORF11770P	Type II	YGGCCR	m5C	Euryarchaeota	(S.-C. Chen et al., 2016)
Methanofollis formosanus ML15	M.MarH5ORF327P	Type II	GCNGC	m5C	Euryarchaeota	(Li et al., 2016)
Methanohalophilus mahii DSM 5219	M.Mma5219II	Type II	AGCT	m4C	Euryarchaeota	(Blow et al., 2016)
Methanohalophilus mahii DSM 5220	M1.Mma5219I	Type I	TCYNNNNNNTCG	m6A	Euryarchaeota	(Morgan et al., 2016)
Methanohalophilus mahii DSM 5221	M2.Mma5219I	Type I	TCYNNNNNNTCG	m4C	Euryarchaeota	(Morgan et al., 2016)
Methanomethylophilus species 1R26	M.Msp1R26ORF4255P	Type II	GCWGC	m5C	Euryarchaeota	http://rebase.neb.com
Methanomethylophilus species 1R26	M.Msp1R26ORF4755P	Type II	GATC	m5C	Euryarchaeota	http://rebase.neb.com
Methanoregula boonei 6A8	M.Mbo6A8ORF1031P	Type I	TGANNNNNNNNTGCT	m6A	Euryarchaeota	http://rebase.neb.com
Methanoregula boonei 6A8	M.Mbo6A8ORF469P	Type II	CCGG	m4C	Euryarchaeota	http://rebase.neb.com
Methanosaeta concilii GP-6	McoGP6ORF996P	Type II	CTGCAG	m6A	Euryarchaeota	http://rebase.neb.com
Methanosarcina mazei LYC	M.MmaLYCORF1026P	Type II	GGWCC	m5C	Euryarchaeota	http://rebase.neb.com
Methanosarcina mazei TMA	M.Mma9314ORF1845P	Type II	GGWCC	m5C	Euryarchaeota	http://rebase.neb.com
Methanosarcina soligelidi SMA-21	M.Mso21ORFBP	Type II	GGCC	m5C	Euryarchaeota	(Alawi et al., 2015)
Methanospirillum hungatei JF-1	M.MhuI	Type II	CTNAG	m4C	Euryarchaeota	(Gunsalus et al., 2016)
Methanospirillum hungatei JF-1	M.MhuII	Type II	GTAC	m4C	Euryarchaeota	(Blow et al., 2016) (Gunsalus et al., 2016)
Methanospirillum hungatei JF-1	M.MhuIII	Type II	GATC	m6A	Euryarchaeota	(Blow et al., 2016)
Methanospirillum hungatei JF-1	M.*Mhu*ORF2537P	Type II	CCCGGG	m4C	Euryarchaeota	(Blow et al., 2016)
Methanospirillum species 1.3.6.1-F.2.7.3	M.Msp273I	Type II	GATC	m6A	Euryarchaeota	http://rebase.neb.com
Methanospirillum species 1.3.6.1-F.2.7.3	M.Msp273II	Type II	GTAC	m4C	Euryarchaeota	http://rebase.neb.com
Methanospirillum species 1.3.6.1-F.2.7.3	M.Msp273III	Type II	AGCT	m4C	Euryarchaeota	http://rebase.neb.com
Methanospirillum species 1.3.6.1-F.2.7.3	M.Msp273IV	Type I	TAGNNNNNGTAG	m6A	Euryarchaeota	http://rebase.neb.com
Methanospirillum species 1.3.6.1-F.2.7.3	M.Msp273V	Type I	TTACNNNNGTC	m6A	Euryarchaeota	http://rebase.neb.com
Methanospirillum species 1.3.6.1-F.2.7.3	M.Msp273VI	Type II	CTNAG	m4C	Euryarchaeota	http://rebase.neb.com
Methanospirillum species 1.3.6.1-F.2.7.3	M2.Msp273VII	Type I	GCAGNNNNNGGC	m4C, m6A	Euryarchaeota	http://rebase.neb.com
Methanotorris igneus Kol 5	M.Mig5ORF1783P	Type II	CCGG	m4C	Euryarchaeota	http://rebase.neb.com
Micrarchaeum acidiphilum ARMAN-2	M.Mac2ORF159P	Type II	CCNGG	m5C	Micrarchaeota	(Dick et al., 2009)
	M.Mac2ORFAP	Type II	CCNGG	m5C	Micrarchaeota	(Dick et al., 2009)
Nanoarchaeota archaeon 7A	Nar7I	Type II	CTGRAG	m6A	Nanoarchaeota	http://rebase.neb.com
Natrialba asiatica	M.Nas12278I	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com
Natrialba asiatica	M.Nas12278II	Type II	CATTC	m6A	Euryarchaeota	http://rebase.neb.com
Natrialba magadii	M.NmaORF4246P	Orphan	GATC	m6A	Euryarchaeota	http://rebase.neb.com
Natrinema pallidum BOL6-1	M.NpaBOL61I	Type II	CATTC	m6A	Euryarchaeota	http://rebase.neb.com
Natrinema pallidum BOL6-1	M.NpaBOL61II	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com

(*continued on next page*)

Table 1 (*continued*)

Organism Name	Restriction Enzyme	Type	Predicted Recognition Sequence	Predicted Methylation Product	Phylum	Reference
Natrinema pellirubrum	M. Npe15624ORF1239P	Type II	GWGCWC	m5C	Euryarchaeota	http://rebase.neb.com
Natrinema versiforme BOL5-4	M.Nve54I	Type II	CATTC	m6A	Euryarchaeota	http://rebase.neb.com
Natrinema versiforme BOL5-4	M.Nve54II	Type II	TCCTCGG	m4C	Euryarchaeota	http://rebase.neb.com
Natrinema versiforme BOL5-4	M.Nve54III	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com
Natronobacterium gregoryi SP2	M.NgrSP2ORF166P	Type II	ACGT	m5C	Euryarchaeota	http://rebase.neb.com
Natronobacterium gregoryi SP2	M.NgrSP2ORF2905P	Type II	RGCGCY	m5C	Euryarchaeota	http://rebase.neb.com
Natronomonas moolapensis 8.8.11	M1. Nmo8811ORF2178P	Type II	GTCGAC	m5C	Euryarchaeota	(Dyall-Smith et al., 2011)
Natronomonas moolapensis 8.8.11	M2. Nmo8811ORF2178P	Type II	GTCGAC	m5C	Euryarchaeota	(Dyall-Smith et al., 2011)
Natronorubrum bangense JCM 10635	M.Nba10635I	Type II	CATTC	m6A	Euryarchaeota	(Xiong et al., 2019)
Nitrosoarchacum limnia BG20	M.NliBG20ORFBP	Type II	GCWGC	m5C	Thaumarchaeota	(Mosier et al., 2012)
Nitrosoarchaeum species isolate BD3	M.NspBD3I	Type II	AGCT	m4C	Thaumarchaeota	(Hiraoka et al., 2019)
	M.NspBD3II	Type II	GATC	m6A	Thaumarchaeota	http://rebase.neb.com
Nitrosomarinus catalina SPOT01	M.Nca01I	Type II	AGCT	m4C	Nitrososphaerota	(Ahlgren et al., 2017)
	M.Nca01II	Type II	GATC	m6A	Nitrososphaerota	(Ahlgren et al., 2017)
Nitrosopelagicus brevis V2	M.NbrV2ORF1407P	Type II	GGCC	m5C	Thaumarchaeota	(Santoro et al., 2015)
Nitrosopumilus species AR2	M.NspAR2ORF7940P	Type II	CCWGG	m5C	Nitrososphaerota	(Park et al., 2012)
Nitrososphaera viennensis EN76	M.NviEN76ORF1230P	Type II	GGWCC	m5C	Thaumarchaeota	http://rebase.neb.com
Nitrososphaera viennensis EN76	M. NviEN76ORF23050P	Type II	GTAC	m5C	Thaumarchaeota	http://rebase.neb.com
Nitrososphaera viennensis PLX03	M. NviPLX03ORF11360P	Type II	GTAC	m5C	Thaumarchaeota	http://rebase.neb.com
Nitrososphaera viennensis PLX03	M.NviPLX03ORF575P	Type II	GGWCC	m5C	Thaumarchaeota	http://rebase.neb.com
Palacococcus ferrophilus DSM 13482	M.Pfe13482I	Type II	GATC	m6A	Euryarchaeota	(Blow et al., 2016)
Palacococcus ferrophilus DSM 13482	M.Pfe13482II	Type I	CGANNNNNNTTTC	m6A	Euryarchaeota	(Blow et al., 2016)
Palacococcus ferrophilus DSM 13482	M.Pfe13482III	Type I	GAAYNNNNNNCTG	m6A	Euryarchaeota	(Blow et al., 2016)
Picrophilus torridus	M.*Pto*ORF585P	Type II	CGCG	m4C	Euryarchaeota	(Fütterer et al., 2004)
Pyrobaculum aerophilum IM2	M.*Pae*IMORF3201P	Type II	GCWGC	m5C	Candidatus Thermoplasmatota	(Fitz-Gibbon et al., 2002)
Pyrobaculum species 186	M.Psp1860ORF1393P	Type II	GCWGC	m5C	Thermoproteota	(Mardanov et al., 2012)
Pyrobaculum species 186	M.Psp1860ORF1413P	Type II	GCWGC	m5C	Thermoproteota	(Mardanov et al., 2012)
Pyrobaculum species WP30	M. PspWP30ORF1023P	Type II	GCWGC	m5C	Thermoproteota	(Jay et al., 2015)
Pyrococcus horikoshii OT3	M.*Pho*I	Type II	GGCC	m4C	Euryarchaeota	(Blow et al., 2016)
Pyrococcus horikoshii OT3	M.PhoII	Type II	GATC	m6A	Euryarchaeota	(Herrero & Thorpe, 2016)
Pyrococcus horikoshii OT3	M.PhoIII	Type II	GCNAGA	m6A	Euryarchaeota	(Cifuentes et al., 2016)
Pyrodictium occultum PL-19	M.PocPL19ORF5910P	Type II	GCWGC	m5C	Thermoproteota	(Utturkar et al., 2016)
Salarchacun species JOR-1	M.SspJOR1I	Type II	CTAG	m4C	Euryarchaeota	http://rebase.neb.com

(*continued on next page*)

Table 1 (*continued*)

Organism Name	Restriction Enzyme	Type	Predicted Recognition Sequence	Predicted Methylation Product	Phylum	Reference
	M.SspJOR1III	Type II	GGWCC	m5C	Euryarchaeota	http://rebase.neb.com
	SspJOR1II	Type II	AGCGANC	m6A	Thermoproteota	http://rebase.neb.com
Staphylothermus hellenicus	M.SmaF1I	Type I	CAYNNNNNNTCA	m6A	Thermoproteota	(Anderson et al., 2009)
Sulfolobus acidocaldarius DSM 639	M.*Sua*I	Type II	GGCC	m4C	Thermoproteota	(L. Chen et al., 2005)
Sulfolobus acidocaldarius DSM 639	M.SuaII	Type II	RGATCY	m5C	Thermoproteota	(L. Chen et al., 2005)
Sulfolobus acidocaldarius DSM 639a	M.Sac639aORF4165P		RGATCY	m5C	Thermoproteota	(L. Chen et al., 2005)
Sulfolobus acidocaldarius DSM 639a	M. SauGG12ORF3145P	Type II	RGATCY	m5C	Thermoproteota	http://rebase.neb.com
Sulfolobus acidocaldarius N8	M.Sac8ORF3145P	Type II	RGATCY	m5C	Thermoproteota	(Mao & Grogan, 2012)
Sulfolobus acidocaldarius NG05B_CO5_07	M. SacNG05ORF6170P	Type II	RGATCY	m5C	Thermoproteota	http://rebase.neb.com
Sulfolobus acidocaldarius Ron12/I	M.Sac12ORF3135P	Type II	RGATCY	m5C	Thermoproteota	http://rebase.neb.com
Sulfolobus acidocaldarius Y14 16-22	M.Sac1622ORFBP	Type II	RGATCY	m5C	Thermoproteota	http://rebase.neb.com
Sulfolobus acidocaldarius Y14 18-5	M.Sac185ORFBP	Type II	RGATCY	m5C	Thermoproteota	http://rebase.neb.com
Sulfolobus acidocaldarius Y14 20-20	M.Sac2020ORFBP	Type II	RGATCY	m5C	Thermoproteota	http://rebase.neb.com
Sulfolobus islandicus Y.N.15.51	M.SisYNORF2183P	Type II	GGCC	m4C	Crenarchaeota	http://rebase.neb.com
Sulfolobus islandicus Y.N.15.51	M.SisYNORF2288P	Type II	ACGGC	–	Crenarchaeota	http://rebase.neb.com
Thaumarchacota archaeon MY3	M.TarMY3I	Type II	AGCT	m4C	Euryarchaeota	http://rebase.neb.com
Thaumarchacota archaeon MY3	M.TarMY3II	Type II	GATC	m6A	Euryarchaeota	http://rebase.neb.com
Thermococcus barophilus CH5	M.TbaCH5ORF1633P	Type II	GGCCG	m5C	Euryarchaeota	(Oger et al., 2016)
Thermococcus kodakaraensis KOD1	TkoI	Type II	GTGAAG (20/18)	m6A	Euryarchaeota	(Fukui et al., 2005)
	TkoII	Type II	TTCAAG (10/8)	m6A	Euryarchaeota	(Fukui et al., 2005)
Thermococcus kodakarensis TS900	M. TkoTS900ORF5857P	Type II	TTCAAG	m6A	Euryarchaeota	http://rebase.neb.com
Thermococcus litoralis	M.*Tli*I	Type II	CTCGAG	m6A	Euryarchaeota	(Gardner et al., 2012)
Thermococcus litoralis	M.TliII	Type II	GCAGG	m6A	Euryarchaeota	(Gardner et al., 2012)
Thermococcus species CL1	M.TspCL1ORF1080P	Type II	CGGCCG	m5C	Euryarchaeota	(Gardner et al., 2012)
Thermococcus species CL1	M.TspCL1ORF1491P	Type II	GCATGC	m5C	Euryarchaeota	(Gardner et al., 2012)
Thermofilum species 1910b	M.Tsp1910ORF3720P	Type II	GGCC	m5C	Thermoproteota	(Dominova et al., 2013)
Thermoplasma acidophilum DSM 1728	M.*Tha*I	Type II	CGCG	m4C	Euryarchaeota	(Ruepp et al., 2000)
Thermoplasma acidophilum DSM 1728	M.ThaII	Type II	GATC	m6A	Euryarchaeota	(Blow et al., 2016) (Koike et al., 2005)
Thermoplasma acidophilum DSM 1728	M.ThaIII	Type II	GANTC	m6A	Euryarchaeota	(Blow et al., 2016) (Ruepp et al., 2000)
Thermoplasma acidophilum DSM 1728	M.ThaIV	Type II	CATG	m6A	Euryarchaeota	(Blow et al., 2016) (Ruepp et al., 2000)
Thermoplasma volcanium GSS1	M.TvoDam	Type II	GATC	m6A	Euryarchaeota	(Kawashima et al., 2000)
Thermoplasma volcanium GSS1	M.TvoI	Type II	CATG	m6A	Euryarchaeota	(Kawashima et al., 2000)

(*continued on next page*)

Table 1 (*continued*)

Organism Name	Restriction Enzyme	Type	Predicted Recognition Sequence	Predicted Methylation Product	Phylum	Reference
Thermoplasma volcanium GSS1	M.TvoII	Type III	CGCC	m4C	Euryarchaeota	(Kawashima et al., 2000)
Thermoplasma volcanium GSS1	M.TvoIII	Type II	GTNAC	m6A	Euryarchaeota	(Kawashima et al., 2000)
Thermoplasma volcanium GSS1	M.TvoIV	Type II	CCWGG	m4C	Euryarchaeota	(Kawashima et al., 2000)
Thermoplasma volcanium GSS1	M.TvoV	Type II	CCSGG	m4C	Euryarchaeota	(Kawashima et al., 2000)
Thermoplasma volcanium GSS1	M.TvoVI	Type II	GANTC	m6A	Euryarchaeota	(Kawashima et al., 2000)
Thermoplasmatales archaeon I-plasma	M.TarIpORFGP	Type II	CGCG	m4C	Euryarchaeota	(Dick et al., 2009)
Thermoproteus neutrophilus V24Sta	M.TneI	Type II	GCWGC	m5C	Crenarchaeota	http://rebase.neb.com
Thermosphaera aggregans	M.TagYORF881P	TypeII	GCWGC	m5C	Thermoproteota	(Spring et al., 2010)

Technology (ONT) platform has also been used to identify restriction enzyme recognition sequences but with limited success [23]. In 2016, a comprehensive study on the whole genome methylomes of numerous bacterial and archaeal species was conducted via SMRT sequencing which deciphered numerous motifs methylated by both orphan methylases and methyltransferases with cognate restriction endonuclease systems [24].

RM systems have been extensively studied in numerous bacterial species such as *E. coli* [25,26], *Salmonella* [27], *Klebsiella pneumoniae* [28], *Citrobacter freundii* [29], *Neisseria gonorrohoeae* [30], *Helicobacter pylori* [31,32] etc. However, in the case of archaeal species, it's a rather unexplored domain. There are only a handful of archaeal species in which the RM systems have been characterized and biochemically tested. This paper aims to review and consolidate the existing knowledge on the RM systems present in archaea shedding light on their cognate methyltransferases and their significance in the ongoing battle between microbial host, their viral adversaries and other auxiliary roles.

1.1. Restriction modification systems characterized in archaea

The research on RM systems has witnessed remarkable progress in recent years which is primarily driven by advancements in sequencing technology. Presently, the REBASE database contains an extensive repository of more than 52,000 RM enzymes [33]. Out of these, 5168 RM enzymes from 615 archaeal species have been elucidated to date (http://rebase.neb.com/rebase/rebase.html). However, the precise recognition sequence and predicted methylated bases remain unrevealed for numerous archaeal species. Table 1 presents a compilation of the archaeal species whose respective recognition sequences and predicted methylated bases have been identified. Maximum RM systems were identified in phylum euryarchaeota. Certain archaeal species such as *Methanocaldococcus jannaschii*, *Methanobrevibacter*, *Methanospirillum*, *Thermoplasma volcanium*, *Sulfolobus acidocaldaris* and *Halobacterium* species were enriched with numerous types of RM systems. Possible reason for the extensive enrichment of RM systems in certain species could be attributed to their intrinsic capacity of natural transformation wherein RM systems aid to impede the entry of incoming DNA.

2. Thermophilic and hyperthermophilic archaeal RM systems

The first RM system in an archaeal organism was identified in the hyperthermophilic archaea *Thermoplasma acidophilum* viz. *Tha*I that thrives at pH 0.7–2 and a temperature of 59 °C. The recognition sequence was determined to be 5′-CG/CG based on the cleavage profile of the plasmid DNA, phage DNA, and genomic DNA upon digestion with *Tha*I. *Tha*I has been classified as a Type II restriction enzyme based on the palindromic recognition sequence. The optimal activity of *Tha*I was observed at a temperature of 60 °C. The genomic DNA of *Thermoplasma* exhibited resistance to *Tha*I digestion, indicating the possibility of methylation by cognate methyltransferase [34]. The identification of *Tha*I provided the first evidence for the presence of RM systems in archaeal organisms and formed the basis for investigating archaeal biology and the evolutionary aspects of RM systems.

This groundbreaking discovery was followed by the isolation of another Type II restriction enzyme, namely *Sua*I, from the thermophilic archaea *Sulfolobus acidocaldaris.* The recognition sequence of this enzyme was determined to be 5′-GG/CC. As expected, the enzyme failed to cleave *Sulfolobus acidocaldaris* DNA*,* suggesting methylation of genomic DNA by cognate methyltransferase. *Sua*I was found to be thermostable, with maximum catalytic activity at 60°C–70 °C and pH 7.5–8 [35]. In a follow-up study, the methylation pattern within the recognition site was elucidated via *in vitro* experiments using isoschizomers of *Sua*I [36]. The results showed that N_4 cytosine was methylated at the inner C residue in 5′-GGCC stretch, resulting in DNA cleavage resistance by the respective endonucleases. Suzuki and Kosarawa created a deletion mutant of *Sua*I and showed that ΔSuaI displayed a loss of restriction endonuclease activity and facilitated the uptake of foreign DNA into the host organism [37].

A Type II RM system was characterized by the thermophilic archaea *Methanobacterium wolfei* viz. *Mwo*I with 5′-GCN5/N2GC as the recognition site. Extracts from *E. coli* transformed with *Mwo*I showed endonuclease activity and the plasmid extracted from the

transformed *E. coli* strain displayed resistance to *Mwo*I cleavage, indicating successful cleavage by REase and plasmid DNA methylation by MTase [38].

Methanothermobacter thermoformicicum THF strain has been found to harbor another Type II RM system viz. *Mth*TI, encoded by plasmid pVF1 with 5′-GG/CC as a recognition sequence [39]. In a follow-up study, Nolling and de Vos further investigated the Z-245 and FTF strains of *M. thermoformicicum* and identified two more RM systems viz. *Mth*ZI and *Mth*FI respectively. The digestion of plasmid and phage DNA using cell free extracts of these two strains revealed 5′-C/TAG as the target recognition sequence.

Another thermostable restriction endonuclease, *Psp*GI, was isolated from *Pyrococcus* sp. strain GI-H. *Psp*GI is an isoschizomer of *Eco*RII and has the recognition sequence 5'-/CCWGG-3 (where W is A or T). This enzyme exhibits optimal activity at temperatures ranging from 65 to 85 °C. Recombinant *Psp*GI expressed in *E. coli* has a half-life of 2 h at 95 °C. Due to its exceptional thermostability, *Psp*GI can potentially be employed for DNA cleavage during DNA amplification processes [40].

Ishwaka et al. used a bioinformatics approach to identify potential restriction endonuclease genes in the genome of *Pyrococcus abyssi* and *Pyrococcus horikoshii*. 32 candidate genes were identified as restriction endonucleases based on their proximity to methyltransferase genes. These genes were cloned and further expressed and out of 32 annotated genes, two candidates exhibited restriction endonuclease activity, viz., *Pho*I (previously identified in *Pyrococcus horikoshii*) and *Pab*I. *Pho*I is known to recognize 5′-GTTAAC, whereas the recognition sequence of *Pab*I has been identified as 5′-GTA/C. The methyltransferase counterpart of *Pab*I (M.*Pab*I) was further investigated in a follow-up study in which *in vitro* methylation assays were set up under different reaction conditions. M.*Pab*I was found to be an adenine methyltransferase that methylates adenine at 5′-GTAC to form 5′-GTm6AC. It displayed remarkable thermostability with maximum activity at 95 °C and pH 5.8 to 6.7. The activation energy and thermodynamic properties of this enzyme were also determined by virtue of its hyperthermophilic properties [41].

Recently, two novel restriction-modification (RM) systems, TkoI and TkoII, have been characterized from the hyperthermophilic archaeon *Thermococcus kodakarensis*, which grows at 80 °C. Zapotek et al. (2019) showed that deletion of either or both of these RM systems abrogated the growth of *T. kodakarensis* and displayed increased transformation efficiency compared to the wild-type strains, suggesting that RM systems act as a barrier to the influx of foreign DNA. TkoI and TkoII were classified as Type IIL RM enzymes that cleave away from the recognition sites at 5′-GTGAAG(N)20/(N)18 and 5′-TTCAAG(N)10/(N)8 respectively. Single-molecule real-time (SMRT) sequencing revealed that TkoI and TkoII methylate adenine residues on one strand of DNA at the GTGA**A**G and TTCA**A**G sites [42]. The discovery of these two new RM systems provides new insights into the mechanisms by which archaea protect their genomes from foreign DNA. The results of this study also have implications for the development of new strategies for the genetic engineering of archaea.

Another study by Gulati et al. showed the characterization of Type I methyltransferase M.PtoI from *Picrophilus torridus*. *P.torridus* is a thermoacidophile that inhabits dry solfataric fields in Japan. M.PtoI was found to methylate adenine at pH 0.7 and temperature 55–60 °C, reflecting the natural habitat of dry solfataric fields wherein *Picrophilus* dwells. Notably, this study marks the first comprehensive examination of a methyltransferase enzyme originating from an organism adapted to such extreme conditions [43].

3. Mesophilic and methanogenic archaeal RM systems

Besides thermophilic archaea, restriction-modification (RM) systems have also been characterized in mesophilic and methanogenic archaea, further expanding the spectrum of organisms in which these systems are observed. Thomm et al. (1988) isolated another restriction endonuclease (REase) from the cellular extracts of *Methanococcus vanielii*, a mesophilic methanogenic archaeon. The REase was named *Mvn*I and recognized DNA sequence at 5′-CG/CG [44]. In contrast to its other hyperthermophilic isoschizomers, *Tha*I and *Fnu*DII are functional at 60 °C., *Mvn*I showed optimum activity at 37 °C, suggesting it as a feasible alternative for laboratory applications.

*Mae*I in *Methanococcus aerolis* and *Mja*I in *Methanocaldococcus jannaschii* were also known to possess CTAG as the recognition site [45]. These findings suggest that CTAG-specific RM systems are more prevalent in archaea compared to bacteria.

Recently, Fomenkov et al. has identified five restriction modification systems encoded by *Methanococcus aeolicus* PL15/Hp by methylome sequencing, homology-based gene annotation and recombinant gene expression [46].

Apart from the above-mentioned RM systems, the crystal structures of the hsdS subunit from *Methanocaldococcus jannaschii* (PDB ID: 1YF2) and the hsdM subunit from *Methanosarchina mazei* (PDB ID: 3KHK) of Type I RM systems have been elucidated in archaeal species [47],https://www.rcsb.org/structure/3KHK). However, there are no investigations on the characterization and biochemical mechanisms of these enzymes to date.

4. Halophilic RM systems

Apart from this, RM systems have also been identified and characterized in archaeal species inhabiting hypersaline environments, such as *Halobacterium* and *Haloferax,* thereby broadening the scope of the environment in which these organisms are observed and studied. Patterson and Pauling examined DNA methylation and RM systems in the halophilic archaeon *Halobacterium salinarum,* also known as *Halobacterium cutirubrum*. The ability of *H. salinarum* to restrict and modify the halophage Hh3 was determined through plating experiments, using *Halobacterium halobium* as the indicator host. It was further shown that *H. salinarum* harbored two RM systems, and the loss of either or both of these systems resulted in the emergence of four distinct RM phenotypes [48]. However, the specific RM systems, their target recognition sites, and methylation patterns were not identified in this study.

A study by Holmes et al. demonstrated that shuttle vectors cloned and propagated in *E. coli,* showed low transformation efficiency upon propagation in *H. volcanii*, indicating the presence of an RM system. Nevertheless, when these vectors were introduced into an

E. coli dam$^-$ mutant, a three-fold increase in transformation efficiency was observed [49]. This suggested the presence of a methylation-dependent restriction enzyme in *H. volcanii* that targets methylated adenine. The whole genome sequencing of *H. volcanii* showed the presence of a hypothetical Type IV restriction endonuclease (REase) gene termed mrr. It was speculated that the mrr gene might play an important role in the reduction of transformation efficiency on methylated DNA within *H. volcanii* [50]. Allers et al. further investigated the hypothesis that the mrr gene encodes a Type IV restriction endonuclease in *H. volcanii* wherein the *mrr* gene was knocked out in *H. volcanii* and Δmrr exhibited high transformation efficiency for both methylated as well as unmethylated plasmids which further confirmed the presence of Type IV restriction endonuclease in *H. volcanii* [51].

Previous findings by Charlebois et al. indicated that the genomic DNA of *H. volcanii* carries methylated cytosine at 5′-CTAG site that is known to be methylated by a putative Type IIG methyltransferase HVO_A0006. DNA extracted from *H. volcanii* was digested with *Xba*I (which is known to cleave the CTAG sequence) and was found to be resistant to digestion [52]. This provided evidence that the *H. volcanii* DNA was methylated at the 5′-CTAG tetranucleotide. The presence of m4C methylation at the 5′-CTAG motif across the genome of *H. volcanii* was further confirmed by a comprehensive analysis of genome-wide methylation via SMRT sequencing [53]. Another type I RM system was identified in *Haloferax volcanii* via SMRT sequencing. The predicted recognition sequence was found to be 5′-GCAB(N)6V**T**GC encoded by the operon HVO_2269–2271 (rmeRMS). The methylation occurs at the third adenine in the upper strand and the adenine complementary to thymine in bold in the lower strand of the recognition sequence [53]. Deletion of these annotated restriction-modification (RM) genes resulted in the absence of methylation at the corresponding DNA sites [54].

4.1. Epigenomic analysis of methylated motifs in archaea

DNA methylation is a widespread epigenetic modification observed across prokaryotes (various bacteria and some archaeal species) and eukaryotes. Methylation is marked by the addition of a methyl group from methyl donor S-adenosyl methionine (SAM) to either adenine or cytosine in the genomic DNA via S_N2 nucleophilic substitution catalyzed by DNA methyltransferases. The resulting modified nucleotides – N^6-methyl adenine (m6A), N^4-methyl-cytosine (m4C), and 5-methyl cytosine (m5C) serve as the bias for the self and foreign DNA in a milieu of restriction-modification system framework. While all three types of methylation are found in prokaryotes, in eukaryotes, m5C is predominantly present in the form of CpG islands [55]. DNA methylation in prokaryotes has been implicated in a myriad of cellular processes including DNA replication, DNA repair, and gene regulation [56–58]. DNA methyltransferases can either work in concert with their cognate restriction endonuclease as a part of the RM system or in solitary as orphan MTases. While the genome-wide methylation status of several bacterial species and eukaryotes has been extensively studied, investigations on the epigenomic landscapes of archaeal genomes is limited.

5. Thermophilic and hyperthermophilic archaeal MTases

Earlier attempts at deciphering the MTase activities in prokaryotes (both bacteria and archaea) utilized digestion sensitivity-based assays. One of the first such studies in archaea characterized the recombinant methyltransferase M.*Pab*I from the hyperthermophilic archaeon *Pyrococcus abyssi* that was shown to methylate 5′-GTAC to generate 5′-GTm6AC [59]. In a similar study, M.*Sua*I from the thermophilic archaeon *Sulfolobus acidocaldarius* DSM639 was shown to carry a methylated 5′-GGm4CC motif in its genome [36].

The emergence of novel sequencing technologies, such as single-molecule real-time (SMRT) sequencing of Pacific Biosciences, has facilitated the characterization of methylated motifs during genome mapping at a large scale in a convenient manner. SMRT sequencing uses fluorescently tagged nucleotides for detecting methylated bases [21]. It involves a SMRT cell with tiny pores called ZMWs that immobilize DNA templates. Each nucleotide is labeled with a distinct fluorescent tag, producing a unique signal upon incorporation. The fluorescence pulse duration is longer for methylated adenine (3–5 times) and methylated cytosine (5–7 times) compared to their unmethylated counterparts. The RS Modification and Motif analysis program analyzes the IPD ratio, which represents the time taken to incorporate a base at a specific position compared to an unmodified base [22]. The Oxford Nanopore Technology (ONT) platform has also been used to identify restriction enzyme recognition sequences but with limited success [23]. In 2016, a comprehensive study on the whole genome methylomes of numerous bacterial and archaeal species was conducted via SMRT sequencing which deciphered numerous motifs methylated by both orphan methylases and methyltransferases with cognate restriction endonuclease systems [24].

The genome of *Palaeococcus ferrophilus* DSM 13482, a hyperthermophile and extremely barophilic archaea that survives at pressures higher than atmospheric pressure was also mapped to look for methylated motifs. A total of five methylated motifs were mapped. Two type I MTases - M.Pfe13482ORFAP and M.Pfe13482ORFFP were correlated to the m6A methylated motifs 5′- C**A**GNNNNNNNR**T**TC and 5′- CG**A**NNNNNNT**T**TC, respectively. A type II MTaseM.Pfe13482I was correlated with m6A methylation at 5′- GATC motif. Other than these, an m6A methylated motif 5′- C**A**ANNNNNT**T**G was linked to multiple MTase candidates within the genome whereas no obvious candidates could be linked with the m4C methylated motif 5′- GTC**C**TC.

Pyrococcus horikoshii OT3 a hyperthermophilic archaeon with optimal growth temperature as high as 98 °C was found to harbor three methylated motifs 5′- G**AT**C (m6A), 5′- GCN**A**GA (m6A), and 5′- G**GC**C (m4C) all correlating to three type II MTases - M.PhoII, M. PhoIII and M.*Pho*I, respectively.

Marine hyperthermophile *Staphylothermus marinus* F1, showed the presence of five methylated motifs in its genome. One of the motifs 5′- C**A**YNNNNNNN**T**CA was correlated to a type I MTase - M.SmaF1I which was responsible for m6A modification. Another m6A modification at the site 5′- G**AT**C was tentatively associated with a type II MTaseM.SmaF1ORF588P. Two of the observed methylated motifs 5′- GGN**A**CB (m6A) and 5′- **C**TCGA**G** (m4C) were correlated to multiple MTase candidates across the genome.

M.*Sau*I, a type II MTasefrom *Sulfolobus acidocaldarius* DSM 639, as previously characterized [36] was correlated to the m4C

methylated motif 5′- G**GC**C. Besides this, another type II MTase- M.SuaII was found to be associated with the m5C methylation at the site 5′- R**GATCY**.

Thermoplasma acidophilum DSM 1728, a thermophilic archaeon was shown to harbor six different methylated motifs in its genome. Of these, four motifs: 5′- G**AT**C, 5′- C**AT**G, 5′- G**ANT**C, and 5′-G**T**N**A**C that exhibited the presence of m6A modification have corresponded assuredly with type II MTases - M.ThaII, M.ThaIV, M.ThaIII, and rather obscurely with M.ThaORF1417P, respectively. A type II MTase – M.*Tha*I was found to be responsible for the m4C methylation at the observed 5′- **C**GC**G** motif. The m4C methylated motif 5′- T**CG**A could not be linked to any obvious MTase candidate from the genome.

A moderately thermophilic archaea *Thermoplasma volcanium* GSS1, was found to harbor at least nine methylated motifs across its genome. Interestingly, only two of these motifs 5′-G**AT**C and 5′-C**AT**G which showed m6A methylation could be certainly associated with type II MTases- M.TvoDam and M.TvoI, respectively. Four other observed motifs – 5′- G**T**N**A**C (m6A), 5′-C**CSG**G (m4C), 5′-CG**C**C (m4C) and 5′-GGR**A**G (m6A) were uncertainly correlated to M.TvoORF1436P (type II), M.*Tvo*ORF1413P (type II), M.TvoORF1464P (type III) and *Tvo*ORF681P (type IIG), respectively. Motifs 5′-GAT**A**C and 5′-TG**A**NTC were found to carry m6A methylation, however, both these motifs were correlated to multiple MTase candidates. m6A methylated motif 5′- DAG**A**TTCW could not be correlated to any obvious MTase candidate within the genome.

The carboxydotrophic hyperthermophilic archaeon *Thermococcus onnurineus* NA1 strains which has a unique ability to grow on carbon monoxide and produce biohydrogen was evolutionary engineered to increase H2 production by successive serial transfers in CO medium for up to 150 transfers [60]. Genome-wide epigenomic analysis by SMRT sequencing of the wild type (2T) and the evolved strain (156T) that witnessed 156 serial transfers revealed the presence of both adenine (m6A) and cytosine (m4C) methylated motifs across the genome. Motif 5′- GT(C/A)(G/T)AC was found in about 95% of the adenine methylation sequences which closely corresponded to the *Acc*I restriction enzyme site. *T. onnurineus* NA1 genome was found to encode for both the *Acc*I restriction enzyme (TON_1382) and its associated methyltransferase (TON_1383). Cytosine methylation was not widely detected across the genome with only 83 and 40 m4C methylated sites in 2T and 156T strains respectively. Although indistinct, the site 5′- Gm4CC was found to be the preferential motif of cytosine methylation at more than 50% of the detected cytosine methylation sites.

A notable study on the genome of an aerobic hyperthermophilic crenarchaeon, *Aeropyrum pernix* K1, which was originally isolated from the coastal solfataras thermal vent in Kodakara-Jima Island in Kyusyu, Japan [61] and is known to grow at an optimum temperature of 95 °C [62] showed the m5C methyltransferase activity of the DNA MTase– M.*Ape*KI at an optimum temperature of 70 °C making it the first highly thermostable DNA m5C methyltransferase to be evaluated experimentally [63]. Usually, DNA Mtases from the organisms thriving at such extremely high temperatures is predicted to carry out either m4C or m6A methylation owing to the susceptibility of m5C DNA modification to heat-induced deamination. This suggests that *A. pernix* K1 presumably harbors a repair mechanism for the high temperature-induced deamination. The study also deciphered the recognition sequence of M.*Ape*KI to be 5′-GC(A/T)GC by methylation activity and bisulfite sequencing (BS-seq). Furthermore, using high-performance liquid chromatography (HPLC) it was found that the M.*Ape*KI adds a methyl group to the second cytosine of 5′-G**C**(A/T)GC.

6. Mesophilic and methanogenic archaeal Mtases

Methanocaldococcus jannaschii DSM 2661, a thermophilic methanogenic archaeon, was found to harbor ten methylated motifs in its genome. Of these, two motifs were tentatively correlated with type I MTases - M.MjaORFCL42P and S.MjaORF1531P responsible for m6A methylation at the sites 5′- CC**A**NNNNNG**T**R and 5′- GC**A**NNNNNNC**TA**, respectively. However, two other m6A methylated sites 5′- G**A**YNNNNNG**T**AA and 5′- CA**A**NNNNNNN**T**GG corresponded to multiple type I MTase candidates. Five methylated motifs were correlated to type II MTases. Of these, two MTases - M.*Mja*I and M.*Mja*V were found to carry out m4C methylation at 5′-**C**TA**G** and 5′-**G**TA**C** sites, respectively. m6A methylation at the sites 5′- G**T**NN**A**C and 5′- G**AT**C were mapped to two type II MTases - M.*Mja*IV and M.*Mja*III, respectively. m5C methylation at 5′- **G**GNC**C** was correlated to M.*Mja*II. A type IIG MTase was also linked to m6A methylation site 5′-CC**A**TC.

Three m4C methylated motifs were found in the genome of another methanogenic archaea *Methanohalophilusmahii* DSM 5219. Of these, two motifs 5′- T**C**YNNNNNNT**C**G and 5′- A**GC**T were correlated to MTases - M1.Mma5219I (type I) and M.Mma5219II (type II), respectively. A tentative type I MTase M2.Mma5219I was predicted to carry out m4C methylation at 5′- T**C**YNNNNNNT**C**G.

An anaerobic methanogenic archaea, *Methanospirillum hungatei* JF-1, commonly found in anaerobic water treatment systems was found to harbor six methylated motifs. Of this M.MhuI, M.MhuII, and M.MhuIV were found to be type II MTases carrying out m4C methylation at the sites – 5′- **C**TNA**G**, 5′- **G**TA**C** and 5′- A**GC**T, respectively. A type IIG MTase - M.MhuIII was correlated to m6A methylated motif - 5′- G**AT**C. Another type IIG MTase MhuORF1081P was speculated to be responsible for m6A methylation at the 5′-GCYYG**A**T motif. However, no certain candidates could be correlated with the observed m4C methylated motif 5′- CCA**C**GK.

7. Halophilic Mtases

Halomicrobium katesii DSM 19301, an extremely halophilic archaea, showed the presence of five methylated motifs in its genome. Two of these motifs 5′- GG**A**YNNNNNNT**G**G and 5′- C**T**CG**A**G exhibited m6A methylation which was mapped to two MTases – a type I MTaseM.Hka19301I and a type II MTase M.Hka19301II, respectively. Two type II MTases - M.Hka19301III and M.Hka19301IV were found to be the definitive candidates responsible for m4C methylation at 5′- **C**TA**G** and 5′- T**C**GC**G**A, respectively. A type IIG MTase Hka19301ORFHP was surmised to be responsible for m6A methylation at the 5′- GATCN**A**G motif.

Another extremely halophilic archaea, *Haloterrigena turkmenica* DSM 5511 originally isolated from sulfate saline soil in Turkmenistan, was also analyzed for the presence of methylated motifs in its genome which revealed the presence of only two

methylated motifs - 5′- **C**TA**G** (m4C) and 5′- C**A**TTC (m6A) both associated with type II MTases - M.HtuI and M.HtuII, respectively.

In another study on archaea, the methylome of aerobic hyperthermoacidophilic crenarchaeon, *Sulfolobus acidocaldarius* DSM 639 was studied using the SMRT technology [64]. Almost 570 previously known m4C methylated motifs 5′- GG**C**C (as identified by Grogan et al., 2003) were found in the genome which validated the stringency of the SMRT sequencing. Interestingly, detailed analyses revealed this motif to be distributed within genes encoding for rRNA and tRNA among many other coding genes. In addition to this, m6A methylation was identified at two different motifs: 5′- AG**A**TCC and 5′- GG**A**TCY containing 5′- G**A**TC as the core motif by SMRT sequencing and dot blot analyses. However, only 19.6% of detected 5′- G**A**TC were fully methylated and hence were believed to not play any apparent role in the protection of genomic DNA.

More recently, genomes (~4Mbp) of two extremely halophilic archaea – *Haloarcula marismortui* and *Haloferax mediterranei*, were sequenced using single-molecule real-time sequencing [65]. Both the genomes were found to harbor m4C methylated 5′-CTAG motif. In *H. marismortui,* about 89% of the 5′- CTAG sites in the genome were methylated and were putatively correlated to a type II –β MTase - M.HmaHMAI (ORF Hma_11,876). Similarly, in *H. mediterranei,* almost 97% of the 5' - CTAG sites were found to be methylated and corresponded to another type II –β MTase - M.Hme33500I (ORF HFX_760). Besides this, genomes of both *H. marismortui* and *H. mediterranei* also harbor a putative type II-α MTase - M.HmaHMAII (ORF Hma_6187) and M.Hme33500II (ORF HFX_3001) responsible for the detected m4C methylation at the motifs: 5′- TCGA**C**GG (85% sites methylated) and 5′- HG**C**W**G**CK (83% sites methylated) respectively. M.HmaHMAII in *H. marismortui* is encoded by pNG600/pHMA155 plasmid (~155 Kbp) and M.Hme33500II in *H. mediterranei* is encoded chromosomally.

In a similar approach, PacBio sequencing was used to study the epigenome of a novel marine Thaumarchaeota, *Candidatus Nitrosomarinus catalina* SPOT01, found in the subsurface Pacific waters off of California [66]. Unlike other Thaumarchaeota, *Ca. Nitrosomarinus catalina* SPOT01 is less tolerant to warm temperatures and is rather better adapted to grow at lower temperatures as low as 10 °C with maximum growth at 23 °C. >99% of the sites: 5′- GATC (m6A) and 5′- AGTC (m4C) were detected to be methylated across the genome. The putative MTase, NMSP_0378 from *N. catalina* was correlated with the m6A methylation at the 5′- GATC site based on its close homology with the MTase - Nmar_1319 from *N. maritimus* which is known to methylate at 5′- GATC [67].

7.1. Orphan methyltransferases in archaea

Orphan MTases are known to function solitarily within the host cells and occur unallied without any cognate restriction endonuclease. Dam, the most profusely studied orphan DNA adenine MTase in bacteria that methylate at second adenine at 5′-G**A**TC motif is known to play a pivotal role in the regulation of DNA replication and repair [68,69]. Another orphan MTase, CcrM found in a gram-negative bacterium *Cauldobacter crescentus*, is known to play a role in cell cycle regulation [70,71]. With several studies on bacteria, reports on orphan MTases within archaeal genomes are rather scarce.

The notable finding on Dam methylase from the archaeal genome dates back to 1986 with Dam^+ phenotype reported in some halophilic and methanogenic archaebacteria species based on the sensitivity to restriction endonucleases – *Dpn*I, *Mbo*I, and *Sau*3AI [72].

In another crucial study, GATC methylation was studied in 21 archaeal species [72]. Of these, 5′-GATC was found to be methylated in the genomes of - *Thermoplasma volcanium, Thermoplasma acidophilum,* and *Pyrococcus* sp. OT3. The genomic DNA of four other archaeal species viz. *Pyrococcus furiosus, Sulfolobus acidocaldarius, Sulfolobus shibatae* and *Sulfolobus solfataricus* showed no GATC methylation. Concomitantly, homologs of *Dam* methylase were also found in these archaea. Six of the archaea under investigation showed the presence of a *Dam* homolog in their genome viz. *Picrophilus torridus*, *Thermoplasma volcanium, Thermoplasma acidophilum, Pyrococcus* sp. OT3, *Methanococcus maripaludis* and *Methanothermococcus jannashii.* Interestingly, while *Picrophilus torridus* harbors a homolog of Dam methylase, it doesn't contain methylated GATC sites in its genome [43].

A recent study on the halophilic archaeal class Halobacteria showed a detailed survey of the presence of RM system genes including orphan DNA methylases across 217 analyzed genomes [73]. The survey revealed that while some orphan MTases were strictly conserved across lineages reflecting their functional significance, the RM system genes exhibited a rather patchy distribution of its presence and absence which was reasoned to be due to regular episodes of horizontal gene transfer and loss of the gene.

Blow et al. conducted a genomic-wide analysis of various archaeal species via SMRT sequencing that revealed the presence of numerous orphan methyltransferases. Orphan methyltransferases, namely M.*Pho*I and M.Pfe113428, which catalyze the methylation of adenine residues at 5′-G**A**TC sites, have been discovered in *Pyrococcus horikoshii* OT3 and *Palaeococcus ferrophilus* DSM 13482. These organisms belong to class Thermococi.

In addition, Dam methylase M.TvoDam and M.ThaII were detected in *Thermoplasma volcanium* GSS1 and *Thermoplasma acidophilum* DSM 1728 which belong to Thermoplasmata class. Moreover, both these Thermoplasmata species were also found to harbor M.TvoI and M.ThaIV which methylate adenine at the 5′-CATG site. m5C methyltransferases were found in *Haloterrigena turkmenica* DSM 5511 and *Halomicrobium katesii* DSM 19301 viz. M.HtuI and M.Hka19301III methylating cytosine at the 5′-CTAG site [24].

7.2. Roles and features of restriction modification systems in archaea

The primary role of the RM system encompasses the defense of bacterial or archaeal species from invading viruses or bacteriophages. RM system recognizes specific DNA sequences of the incoming viruses and bacteriophages and restriction endonuclease ~~cleaves them off~~ degrades the invading DNA, thereby preventing the replication of the virus in the host cell. Although the knowledge of bacterial viruses has been known for a long time, research on archaeal viruses began in the 1980s. The first archaeal virus was identified in *Sulfolobus* viz. *Sulfolobus spindle-shaped virus* 1 (SSV1) [74]. The pioneering research laid the foundation for the

advancement of the fascinating domain of archaeal virology. Currently, archaeal viruses have been classified into 20 families, however taxonomic classification of various groups needs to be addressed [75]. Most of the archaeal viruses are known to infect hyperhalophiles and hyperthermophilies of phyla euryachaeota and crenarchaeota [76]. However, a number of viruses have also been reported to be isolated from methanogens and ammonia oxidizing thaumarcheotal species [77,78]. Unlike bacterial and eukaryotic viruses which have either DNA or RNA as genetic material, all the archaeal viruses characterized to date were found to harbor DNA genome [79].

Archaeal viruses are known to exhibit diverse morphologies and genetic characteristics, distinct from the currently known group of viruses and thus represents the most enigmatic and intriguing group in the world of virosphere. They are known to occur in a wide array of shapes such as spindle-shaped which includes *Sulfolobus spindle shape virus* 1(SSV1), *Sulfolobus spindle-shaped virus* 6 (SSV6), *Pyrococcus abyssi virus* 1(PAV1), *Thermococcus prieurii virus* 1 (TPV1). *Sulfolobus neozealandicus* (SNDV), *Aeropyrum pernix ovoid virus* 1 (APOV1) possess droplet-shaped morphology. Examples of spherical-shaped archaeal viruses are *Pyrobaculum spherical virus* (PSV) and *Haloarcula hispanica icosahedral virus* 2 (HHIV 2). Linear shape viruses include *Sulfolobus islandicus filamentous virus* (SIFV), *Acidianus filamentous virus* 1 (AFV1) [79].

However, the current inventory of identified archaeal viruses is significantly less as compared to the known bacterial and eukaryotic viruses. One possible reason for the comparatively small number of archaeal viruses could be attributed to the challenges associated with culturing of their host species [76]. However, with the advances in the field of metagenomics, there has been an increase in the number of characterized archaeal viruses in the past decade.

Besides undergoing horizontal gene transfer between different archaeal and bacterial species, RM systems also perform a key role in impeding the ingress of exogenous DNA, encompassing viral DNA and various mobile genetic elements thereby protecting the archaeal genome from potentially harmful foreign DNA. The diverse repertoire of restriction enzymes creates a barrier for gene transfer between different strains and species thereby retaining the genetic traits of the population and maintaining speciation. Additionally, the dynamic nature of the RM system leads to the generation of novel restriction and modification sites that further enhance genetic diversity. The appearance of a novel RM system occurs through a mutation in the DNA binding domain of the endonuclease and methyltransferase. Type I RM systems have a DNA recognition module in a separate subunit viz. specificity (S) subunit that interacts with both MTase and REase for recognition of DNA. On the other hand, Type III enzymes recognize DNA via MTase. Type II enzymes have recognition domains in REase and MTase which functions as individual units barring Type IIG that have both enzyme functions in one protein. Specificity change is more probable in Type I, Type III, and Type IIG enzymes in comparison to Type II as a mutation in the S subunit is sufficient to induce the specificity change in the former whereas the latter requires two concurrent mutations in Mtase and REase [80]. These mechanisms of mutations in the specificity subunit enable bacteria and archaea to respond to evolving challenges posed by viruses and phages by changing the recognition sites of their restriction enzymes. The ability of Type I and Type III RM systems to change specificity quickly makes them an effective tool for population maintaining heterogeneity.

Interestingly, in certain archaeal species RM systems were found to play miscellaneous role. Patterson et al. showed the association of RM genes with gas vacuolation in *Halobacterium*. To investigate the correlation of RM system with gas vacuolation, RM systems were eliminated in *Halobacterium cutirubrum* substrain 2 and 3 by irradiation with UV light and it was observed that the restriction negative (Res^{-}) strains lost the property of gas vacuolation [48].

DNA methylation is known to have an impact on the expression of genes. Orphan MTase genes that lack cognate REase are known to be associated with gene regulation and replication. Classical examples of such genes are Dam Mtase in *E.coli* which methylates GATC and Ccr MTase which methylates GANTC in *Caulobacter crescentus* [81]. In *E.coli*, methylation of GATC is known to play a pivotal role in mismatch repair, gene expression, and DNA replication [81]. Koike et al. showed that Dam methylase is not universally present in all archaeal species. This group further elucidated that the upstream region of these dam methylase genes in the archaeal species was found to have a consensus sequence for the binding site of transcription factor FRPP which is known to play an essential role in the regulation of transcription in archaea [82].

7.3. Evolution of RM systems

The presence of RM systems in archaea could also be attributed to horizontal gene transfer. Evolutionary studies involving sequence comparisons and phylogenetic tree analysis suggest that RM genes have undergone horizontal gene transfer between bacteria and archaea based on the sequence homology, codon usage, and GC content [83]. The M.*Pab*I methyltransferase from *Pyrococcus abyssi* exhibited significant sequence similarity with the methyltransferase of *Hin*dII and other bacterial species highlighting the potential exchange of genetic material between archaea and bacteria. Furthermore, it was observed that the GC content of M.*Pab*I was lower than the rest of the ORFs in *P.abyssi* genome further strengthening the concept of HGT [41]. The amino acid sequence of the *Methanothermobacter thermoformicicum* methyltransferase (*Mth*TI) and restriction endonuclease showed significant sequence similarity to *Neisseria gonorrhoeae* RM genes further supporting the notion of migration of genes by HGT between archaea and bacteria [84]. Similarly, there are reports demonstrating sequence similarity in plasmids of *Thermococcale* and *Methanococcale* [85]. In certain archaeal species such as *Methanocaldococcus janachii* and *Methanobacterim thermoformicicium* RM systems are known to reside in plasmids [83]. Genomic analysis by Oliveira et al. showed a compelling correlation between the prevalence of MGE and RM in the genome of bacteria and archaea [86]. The findings indicate that though RM systems are seldom encoded by plasmids, they still disseminate via mobile genetic elements (MGE). RM system is also recognized for its pivotal role in genomic rearrangement. RM system of *Pyrococcus* exhibits polymorphism as evidenced by the presence of homologous segments of RM genes at different loci in the chromosome [87]. This strongly recommends the migration or transposition of DNA segments containing RM genes at different positions in the genome. The mobility system via horizontal gene transfer (HGT) or transposition within the genome supports the hypothesis that RM systems can cause evolutionary changes in the genome and thus can be hypothesized as an important driver of

evolution.

RM systems are also known to have an impact on oligonucleotide composition in the genome of bacteria and archaea. The occurrence of recognition sites of the RM system occurs at a lower frequency than statistically expected [84]. This avoidance of RM sites correlates with the prevention of self-digestion of the genomic DNA of the host by the resident restriction endonuclease. RM genes are found to be tightly linked to each other. For instance, in *Picrophilus torridus* restriction (PTO0076), modification (PTO0078) and specificity (PTO0077) genes are arranged in the linear order [43]. This tight linkage ensures the simultaneous loss of RM genes and is particularly critical for their maintenance by the virtue of post segregational killing [84].

RM systems are known to act as selfish mobile elements. Numerous studies have shown the loss of the RM system can negatively impact cell viability. Another feature supporting this hypothesis is post-segregational killing wherein a decrease in copy number of RM genes triggers cell death [84]. This ensures stable maintenance of RM genes in bacterial and archaeal populations. The mechanism of host killing has been demonstrated in various Type II systems such as *Eco*RI [88], *Eco*RV [89], *Bsp*6I [90] etc. While type I RM systems have not been directly linked to post-segregational killing as reported in the case of EcoR124I wherein the viability of host cells was unaltered after loss of RM genes [91]. In the case of EcoKI, it has been demonstrated that the loss of the RM system can be compensated or substituted by alleles that provide different specificities [92].

RM systems also exhibit superinfection exclusion wherein RM systems compete with each other in several ways. The possible reason could be the attack of resident RM system in the host cell on the incoming RM system due to lack of appropriate methylation. However, if the incoming RM system may have proper methylation site that makes it recalcitrant to cleavage by the resident RM system, still the establishment of external RM system is aborted. It occurs specially in the RM systems that utilizes same regulatory system for establishment. RM systems are also known to compete with each other on the basis of the recognition sequence [84].

Thus, the facts presented above highlights the fact that RM systems are also known to act as mobile genetic elements that are involved in genome arrangement and can be considered as the driving force of evolution and speciation.

8. Conclusion

The role of the RM system in archaea extends beyond the defense mechanism against foreign entities. These systems play a key role in maintaining genetic diversity, evolution, genome rearrangement, and horizontal gene transfer. The dynamic nature of RM systems with their ability to alter the specificity domain provides a further route for the diversification of these enzymes. In conclusion, the advancement of metagenomics coupled with the latest sequencing technologies offers a fertile ground for exploring the diversity of these molecular machines at a vast scale in archaea.

Future directions

Knowledge on archaeal RM system is still fragmentary. Expansion in the inventory of novel RM systems in archaea will provide further insights into the molecular mechanisms of RM enzymes, virus-host interactions, and their role in archaeal biology. With the advancement in sequencing technology, it can be speculated that many more RM enzymes would be added to the library. Investigation of crystal structures of RM enzymes in archaeal species would provide a comprehensive understanding of the molecular mechanism of DNA recognition and modification and further pave the way for the synthesis of novel recombinant RM enzymes with potential applications in biotechnology and genetic engineering. Understanding the contribution of RM systems in the survival of archaea in extreme and inhabitable environments may also offer a further clue for the adaptation of such organisms in the environment where life ceases to exist.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or not-for-profit sector.

Data availability statement

No data was used for the research described in the article.

CRediT authorship contribution statement

Pallavi Gulati: Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Formal analysis, Data curation. **Ashish Singh:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Sandeep Patra:** Methodology, Formal analysis. **Shreyas Bhat:** Methodology, Formal analysis. **Anil Verma:** Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank all the members of our team for the constructive discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27382.

References

[1] W. Arber, D. Dussoix, Host specificity of DNA produced by Escherichia coli, J. Mol. Biol. 5 (1) (1962 Jul) 18–36.
[2] S.E. Luria, M.L. Human, A nonhereditary, host-induced variation of bacterial viruses, J. Bacteriol. 64 (4) (1952 Oct) 557–569.
[3] S.H. Chan, B.L. Stoddard, S y Xu, Natural and engineered nicking endonucleases–from cleavage mechanism to engineering of strand-specificity, Nucleic Acids Res. 39 (1) (2011 Jan 1) 1–18.
[4] R. Sitaraman, The role of DNA restriction-modification systems in the biology of Bacillus anthracis, Front. Microbiol. (2016 Jan 22) 7.
[5] G.G. Wilson, N.E. Murray, Restriction and modification systems, Annu. Rev. Genet. 25 (1) (1991 Dec) 585–627.
[6] D.T.F. Dryden, Nucleoside triphosphate-dependent restriction enzymes, Nucleic Acids Res. 29 (18) (2001 Sep 15) 3728–3741.
[7] W.A.M. Loenen, D.T.F. Dryden, E.A. Raleigh, G.G. Wilson, Type I restriction enzymes and their relatives, Nucleic Acids Res. 42 (1) (2014 Jan 1) 20–44.
[8] A. Pingoud, A. Jeltsch, Structure and function of type II restriction endonucleases, Nucleic Acids Res. 29 (18) (2001 Sep 15) 3705–3727.
[9] A. Pingoud, M. Fuxreiter, V. Pingoud, W. Wende, Type II restriction endonucleases: structure and mechanism, Cell. Mol. Life Sci. 62 (6) (2005 Mar) 685–707.
[10] G. Wilson, Type II restriction—modification systems, Trends Genet. 4 (11) (1988 Nov) 314–318.
[11] P. Bist, S. Sistla, V. Krishnamurthy, A. Acharya, B. Chandrakala, D.N. Rao, S-adenosyl-l-methionine is required for DNA cleavage by type III restriction enzymes, J. Mol. Biol. 310 (1) (2001 Jun) 93–109.
[12] I.A. Murray, Y.A. Luyten, A. Fomenkov, N. Dai, I.R. Corrêa, W.G. Farmerie, et al., Structural and functional diversity among Type III restriction-modification systems that confer host DNA protection via methylation of the N4 atom of cytosine, PLoS One 16 (7) (2021) e0253267.
[13] N.K. Raghavendra, Mechanistic insights into type III restriction enzymes, Front. Biosci. 17 (1) (2012) 1094.
[14] D.N. Rao, D.T.F. Dryden, S. Bheemanaik, Type III restriction-modification enzymes: a historical perspective, Nucleic Acids Res. 42 (1) (2014 Jan 1) 45–55.
[15] T. Dimitriu, M.D. Szczelkun, E.R. Westra, Evolutionary ecology and interplay of prokaryotic innate and adaptive immune systems, Curr. Biol. 30 (19) (2020 Oct) R1189–R1202.
[16] S. Sistla, D.N. Rao, S-Adenosyl-L-methionine–Dependent restriction enzymes, Crit. Rev. Biochem. Mol. Biol. 39 (1) (2004 Jan 29) 1–19.
[17] W. Siwek, H. Czapinska, M. Bochtler, J.M. Bujnicki, K. Skowronek, Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI, Nucleic Acids Res. 40 (15) (2012 Aug) 7563–7572.
[18] N.C. Kan, J.A. Lautenberger, M.H. Edgell, C.A. Hutchison, The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes, J. Mol. Biol. 130 (2) (1979 May) 191–209.
[19] J.V. Ravetch, K. Horiuchi, N.D. Zinder, Nucleotide sequence of the recognition site for the restriction-modification enzyme of Escherichia coli B, Proc. Natl. Acad. Sci. USA 75 (5) (1978 May) 2266–2270.
[20] V. Nagaraja, M. Stieger, C. Nager, S.M. Hadi, T.A. Bickle, The nucleotide sequence recognised by the Escherichia coli D type I restriction and modification enzyme, Nucleic Acids Res. 13 (2) (1985) 389–399.
[21] S. Ardui, A. Ameur, J.R. Vermeesch, M.S. Hestand, Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics, Nucleic Acids Res. 46 (5) (2018 Mar 16) 2159–2168.
[22] A. Rhoads, K.F. Au, PacBio sequencing and its applications, Dev. Reprod. Biol. 13 (5) (2015 Oct) 278–289.
[23] J.T. Simpson, R.E. Workman, P.C. Zuzarte, M. David, L.J. Dursi, W. Timp, Detecting DNA cytosine methylation using nanopore sequencing, Nat. Methods 14 (4) (2017 Apr 20) 407–410.
[24] M.J. Blow, T.A. Clark, C.G. Daum, A.M. Deutschbauer, A. Fomenkov, R. Fries, et al., The epigenomic landscape of prokaryotes, PLoS Genet. 12 (2) (2016 Feb 12) e1005854.
[25] A. Piekarowicz, J.D. Goguen, The DNA sequence recognized by the EcoDXX1 restriction endonuclease, Eur. J. Biochem. 154 (2) (1986 Jan) 295–298.
[26] C. Price, J.C. Shepherd, T.A. Bickle, DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different DNA specificities, EMBO J. 6 (5) (1987 May) 1493–1497.
[27] V. Nagaraja, J.C.W. Shepherd, T. Pripfl, T.A. Bickle, Two type I restriction enzymes from Salmonella species, J. Mol. Biol. 182 (4) (1985 Apr) 579–587.
[28] V. Chin, KpnBI is the prototype of a new family (IE) of bacterial type I restriction-modification system, Nucleic Acids Res. 32 (18) (2004 Oct 11) e138–e138.
[29] P. Kannan, G.M. Cowan, A.S. Daniel, A.A.F. Gann, N.E. Murray, Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes, J. Mol. Biol. 209 (3) (1989 Oct) 335–344.
[30] M. Adamczyk-Poplawska, P. Bacal, A. Mrozek, N. Matczynska, A. Piekarowicz, A. Kwiatek, Phase-variable Type I methyltransferase M.NgoAV from Neisseria gonorrhoeae FA1090 regulates phasevarion expression and gonococcal phenotype, Front. Microbiol. (2022 Oct 4) 13.
[31] Y. Furuta, H. Namba-Fukuyo, T.F. Shibata, T. Nishiyama, S. Shigenobu, Y. Suzuki, et al., Methylome diversification through changes in DNA methyltransferase sequence specificity, PLoS Genet. 10 (4) (2014 Apr 10) e1004272.
[32] J. Krebes, R.D. Morgan, B. Bunk, C. Spröer, K. Luong, R. Parusel, et al., The complex methylome of the human gastric pathogen *Helicobacter pylori*, Nucleic Acids Res. 42 (4) (2014 Feb 1) 2415–2432.
[33] R.J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE: a database for DNA restriction and modification: enzymes, genes and genomes, Nucleic Acids Res. 51 (D1) (2023 Jan 6) D629–D630.
[34] D.J. McConnell, D.G. Searcy, J.G. Sutcliffe, A restriction enzyme Tha I from the thermophilic mycoplasma Thermoplasma acidophilum, Nucleic Acids Res. 5 (6) (1978) 1729–1739.
[35] D.A. Prangishvili, R.P. Vashakidze, M.G. Chelidze, Gabriadze Iyu, A restriction endonuclease *Sua* I from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*, FEBS Lett. 192 (1) (1985 Nov 11) 57–60.
[36] D.W. Grogan, Cytosine methylation by the *sua* I restriction-modification system: implications for genetic fidelity in a hyperthermophilic archaeon, J. Bacteriol. 185 (15) (2003 Aug) 4657–4661.
[37] S. Suzuki, N. Kurosawa, Disruption of the gene encoding restriction endonuclease SuaI and development of a host–vector system for the thermoacidophilic archaeon Sulfolobus acidocaldarius, Extremophiles 20 (2) (2016 Mar 20) 139–148.
[38] K.D. Lunnen, R.D. Morgan, C.J. Timan, J.A. Krzycki, J.N. Reeve, G.G. Wilson, Characterization and cloning of MwoI (GCN7GC), a new type-II restriction-modification system from Methanobacterium wolfei, Gene 77 (1) (1989 Apr 15) 11–19.
[39] J. Nölling, WM de Vos, Identification of the CTAG-recognizing restriction-modification systems *Mth* ZI and *Mth* FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene, Nucleic Acids Res. 20 (19) (1992) 5047–5052.

[40] R. Morgan, J ping Xiao, S yong Xu, Characterization of an extremely thermostable restriction enzyme, *psp* GI, from a *Pyrococcus* strain and cloning of the *psp* GI restriction-modification system in *Escherichia coli*, Appl. Environ. Microbiol. 64 (10) (1998 Oct) 3669–3673.
[41] K. Ishikawa, Discovery of a novel restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-GTA/C) from the hyperthermophilic archaeon Pyrococcus abyssi, Nucleic Acids Res. 33 (13) (2005 Jul 21) e112–e112.
[42] K.M. Zatopek, B.W. Burkhart, R.D. Morgan, A.M. Gehring, K.A. Scott, T.J. Santangelo, et al., The hyperthermophilic restriction-modification systems of Thermococcus kodakarensis protect genome integrity, Front. Microbiol. (2021 May 20) 12.
[43] P. Gulati, A. Singh, M. Goel, S. Saha, The extremophile Picrophilus torridus carries a DNA adenine methylase M.PtoI that is part of a Type I restriction-modification system, Front. Microbiol. (2023 Mar 15) 14.
[44] M. Thomm, MvnI: a restriction enzyme in the archaebacterium Methanococcus vannielii, FEMS Microbiol. Lett. 52 (3) (1988 Aug) 229–233.
[45] J. Nölling, W.M. de Vos, Characterization of the archaeal, plasmid-encoded type II restriction-modification system MthTI from Methanobacterium thermoformicicum THF: homology to the bacterial NgoPII system from Neisseria gonorrhoeae, J. Bacteriol. 174 (17) (1992 Sep) 5719–5726.
[46] A. Fomenkov, P. Weigele, C. McClung, C. Madinger, R.J. Roberts, Complete genome assembly and methylome dissection of Methanococcus aeolicus PL15/Hp, Front. Microbiol. (2023 Apr 5) 14.
[47] J.S. Kim, A. DeGiovanni, J. Jancarik, P.D. Adams, H. Yokota, R. Kim, et al., Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications, Proc. Natl. Acad. Sci. USA 102 (9) (2005 Mar 22) 3248–3253.
[48] N.H. Patterson, C. Pauling, Evidence for two restriction-modification systems in Halobacterium cutirubrum, J. Bacteriol. 163 (2) (1985 Aug) 783–784.
[49] M.L. Holmes, S.D. Nuttall, M.L. Dyall-Smith, Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase, J. Bacteriol. 173 (12) (1991 Jun) 3807–3813.
[50] A.L. Hartman, C. Norais, J.H. Badger, S. Delmas, S. Haldenby, R. Madupu, et al., The complete genome sequence of Haloferax volcanii DS2, a model archaeon, PLoS One 5 (3) (2010 Mar 19) e9605.
[51] T. Allers, S. Barak, S. Liddell, K. Wardell, M. Mevarech, Improved strains and plasmid vectors for conditional overexpression of his-tagged proteins in *Haloferax volcanii*, Appl. Environ. Microbiol. 76 (6) (2010 Mar 15) 1759–1769.
[52] R.L. Charlebois, W.L. Lam, S.W. Cline, W.F. Doolittle, Characterization of pHV2 from Halobacterium volcanii and its use in demonstrating transformation of an archaebacterium, Proc. Natl. Acad. Sci. USA 84 (23) (1987 Dec) 8530–8534.
[53] M. Ouellette, J. Gogarten, J. Lajoie, A. Makkay, R. Papke, Characterizing the DNA methyltransferases of Haloferax volcanii via bioinformatics, gene deletion, and SMRT sequencing, Genes 9 (3) (2018 Feb 27) 129.
[54] M. Ouellette, L. Jackson, S. Chimileski, R.T. Papke, Genome-wide DNA methylation analysis of Haloferax volcanii H26 and identification of DNA methyltransferase related PD-(D/E)XK nuclease family protein HVO_A0006, Front. Microbiol. (2015 Apr 8) 6.
[55] J. Korlach, S.W. Turner, Going beyond five bases in DNA sequencing, Curr. Opin. Struct. Biol. 22 (3) (2012 Jun) 251–261.
[56] J. Casadesús, in: Bacterial DNA Methylation and Methylomes, 2016, pp. 35–61.
[57] L. Feng, J. Lou, in: DNA Methylation Analysis, 2019, pp. 181–227.
[58] M.A. Sánchez-Romero, I. Cota, J. Casadesús, DNA methylation in bacteria: from the methyl group to the methylome, Curr. Opin. Microbiol. 25 (2015 Jun) 9–16.
[59] M. Watanabe, H. Yuzawa, N. Handa, I. Kobayashi, Hyperthermophilic DNA methyltransferase M.PabI from the archaeon *Pyrococcus abyssi*, Appl. Environ. Microbiol. 72 (8) (2006 Aug) 5367–5375.
[60] S.H. Lee, M.S. Kim, S.G. Kang, H.S. Lee, Biohydrogen production of obligate anaerobic archaeon Thermococcus onnurineus NA1 under oxic conditions via overexpression of frhAGB-encoding hydrogenase genes, Biotechnol. Biofuels 12 (1) (2019 Dec 8) 24.
[61] Y. Sako, N. Nomura, A. Uchida, Y. Ishida, H. Morii, Y. Koga, et al., Aeropyrum pernix gen. Nov., sp. nov., a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100 C, Int. J. Syst. Bacteriol. 46 (4) (1996 Oct 1) 1070–1077.
[62] Y. Kawarabayasi, Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1, DNA Res. 6 (2) (1999 Jan 1) 83–101.
[63] M. Hayashi, K. Sugahara, A. Yamamura, Y. Iida, Evaluation of the properties of the DNA methyltransferase from Aeropyrum pernix K1, Microbiol. Spectr. 9 (2) (2021 Oct 31).
[64] M. Couturier, A.C. Lindås, The DNA methylome of the hyperthermoacidophilic crenarchaeon Sulfolobus acidocaldarius, Front. Microbiol. (2018 Feb 8) 9.
[65] S. DasSarma, A. Fomenkov, S.L. DasSarma, T. Vincze, P. DasSarma, R.J. Roberts, Methylomes of two extremely halophilic *archaea* species, Haloarcula marismortui and Haloferax mediterranei, Microbiol Resour Announc 8 (27) (2019 Jul 3).
[66] N.A. Ahlgren, Y. Chen, D.M. Needham, A.E. Parada, R. Sachdeva, V. Trinh, et al., Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems, Environ. Microbiol. 19 (6) (2017 Jun 26) 2434–2452.
[67] R.J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE—a database for DNA restriction and modification: enzymes, genes and genomes, Nucleic Acids Res. 43 (D1) (2015 Jan 28) D298–D299.
[68] S. Adhikari, P.D. Curtis, DNA methyltransferases and epigenetic regulation in bacteria, FEMS Microbiol. Rev. 40 (5) (2016 Sep) 575–591.
[69] J. Collier, Epigenetic regulation of the bacterial cell cycle, Curr. Opin. Microbiol. 12 (6) (2009 Dec) 722–729.
[70] D. Gonzalez, J. Collier, <scp>DNA</scp> methylation by <scp>CcrM</scp> activates the transcription of two genes required for the division of <scp>C</scp> aulobacter crescentus, Mol. Microbiol. 88 (1) (2013 Apr 11) 203–218.
[71] D. Gonzalez, J.B. Kozdon, H.H. McAdams, L. Shapiro, J. Collier, The functions of DNA methylation by CcrM in *Caulobacter crescentus* : a global approach, Nucleic Acids Res. 42 (6) (2014 Apr 1) 3720–3735.
[72] D. Lodwick, H.N.M. Ross, J.E. Harris, J.W. Almond, W.D. Grant, Dam methylation in the archaebacteria, Microbiology (N Y). 132 (11) (1986 Nov 1) 3055–3059.
[73] M.S. Fullmer, M. Ouellette, A.S. Louyakis, R.T. Papke, J.P. Gogarten, The patchy distribution of restriction–modification system genes and the conservation of orphan methyltransferases in Halobacteria, Genes 10 (3) (2019 Mar 19) 233.
[74] C. Schleper, K. Kubo, W. Zillig, The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA, Proc. Natl. Acad. Sci. USA 89 (16) (1992 Aug 15) 7645–7649.
[75] E.M. Adriaenssens, M.B. Sullivan, P. Knezevic, L.J. van Zyl, B.L. Sarkar, B.E. Dutilh, et al., Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV bacterial and archaeal viruses subcommittee, Arch. Virol. 165 (5) (2020 May 11) 1253–1260.
[76] J. Munson-McGee, J. Snyder, M. Young, Archaeal viruses from high-temperature environments, Genes 9 (3) (2018 Feb 27) 128.
[77] J.G. Kim, S.J. Kim, V. Cvirkaite-Krupovic, W.J. Yu, J.H. Gwak, M. López-Pérez, et al., Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea, Proc. Natl. Acad. Sci. USA 116 (31) (2019 Jul 30) 15645–15650.
[78] V.Q.H. Ngo, F. Enault, C. Midoux, M. Mariadassou, O. Chapleur, L. Mazéas, et al., Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics, Environ. Microbiol. 24 (10) (2022 Oct 18) 4853–4868.
[79] D.P. Baquero, Y. Liu, F. Wang, E.H. Egelman, D. Prangishvili, M. Krupovic, in: Structure and Assembly of Archaeal Viruses, 2020, pp. 127–164.
[80] A.S. Ershova, I.S. Rusinov, S.A. Spirin, A.S. Karyagina, A.V. Alexeevski, Role of restriction-modification systems in prokaryotic evolution and ecology, Biochemistry (Moscow) 80 (10) (2015 Oct 17) 1373–1386.
[81] H.J. Seong, S.W. Han, W.J. Sul, Prokaryotic DNA methylation and its functional roles, J. Microbiol. 59 (3) (2021 Mar 23) 242–248.
[82] H. Koike, K. Yokoyama, T. Kawashima, T. Yamasaki, S ichi Makino, L. Clowney, et al., GATC Methylation by Dam methylase in archaea: its roles and possible transcription regulation by an FFRP, Proceedings of the Japan Academy, Series B. 81 (7) (2005) 278–290.
[83] J. Nölling, WM de Vos, Identification of the CTAG-recognizing restriction-modification systems *Mth* ZI and *Mth* FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene, Nucleic Acids Res. 20 (19) (1992) 5047–5052.
[84] I. Kobayashi, Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution, Nucleic Acids Res. 29 (18) (2001 Sep 15) 3742–3756.
[85] H. Wang, N. Peng, S.A. Shah, L. Huang, Q. She, Archaeal extrachromosomal genetic elements, Microbiol. Mol. Biol. Rev. 79 (1) (2015 Mar) 117–152.
[86] P.H. Oliveira, M. Touchon, E.P.C. Rocha, The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts, Nucleic Acids Res. 42 (16) (2014 Sep 15) 10618–10631.

[87] A. Chinen, I. Uchiyama, I. Kobayashi, Comparison between Pyrococcus horikoshii and Pyrococcus abyssi genome sequences reveals linkage of restriction–modification genes with large genome polymorphisms, Gene 259 (1–2) (2000 Dec) 109–121.
[88] K. Kusano, T. Naito, N. Handa, I. Kobayashi, Restriction-modification systems as genomic parasites in competition for specific sequences, Proc. Natl. Acad. Sci. USA 92 (24) (1995 Nov 21) 11095–11099.
[89] Y. Nakayama, I. Kobayashi, Restriction-modification gene complexes as selfish gene entities: roles of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion, Proc. Natl. Acad. Sci. USA 95 (11) (1998 May 26) 6442–6447.
[90] S. Kulakauskas, A. Lubys, S.D. Ehrlich, DNA restriction-modification systems mediate plasmid maintenance, J. Bacteriol. 177 (12) (1995 Jun) 3451–3454.
[91] E.M. Kulik, T.A. Bickle, Regulation of the activity of the type ICEcoR124I restriction enzyme, J. Mol. Biol. 264 (5) (1996 Dec) 891–906.
[92] M. O'Neill, A. Chen, N.E. Murray, The restriction–modification genes of *Escherichia coli* K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities, Proc. Natl. Acad. Sci. USA 94 (26) (1997 Dec 23) 14596–14601.