frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 09 November 2023
DOI 10.3389/fmicb.2023.1241399

OPEN ACCESS

EDITED BY
Marleen van Wolferen,
University of Freiburg, Germany

REVIEWED BY
Tamara Basta,
UMR9198 Institut de Biologie Intégrative de la Cellule (I2BC), France
Nuno Machado,
Oxford Nanopore Technologies, United Kingdom
Shamphavi Sivabalasarma,
University of Freiburg, Germany,
in collaboration with reviewer NM

*CORRESPONDENCE
Dina Grohmann
✉ dina.grohmann@ur.de
Felix Grünberger
✉ felix.gruenberger@ur.de

†These authors have contributed equally to this work and share first authorship

RECEIVED 16 June 2023
ACCEPTED 23 October 2023
PUBLISHED 09 November 2023

CITATION
Stöckl R, Nißl L, Reichelt R, Rachel R, Grohmann D and Grünberger F (2023) The transcriptional regulator EarA and intergenic terminator sequences modulate archaellation in *Pyrococcus furiosus*. *Front. Microbiol.* 14:1241399. doi: 10.3389/fmicb.2023.1241399

The transcriptional regulator EarA and intergenic terminator sequences modulate archaellation in *Pyrococcus furiosus*

Richard Stöckl[1†], Laura Nißl[1†], Robert Reichelt[1], Reinhard Rachel[2], Dina Grohmann[1*] and Felix Grünberger[1*]

[1]Institute of Microbiology and Archaea Centre, Faculty for Biology and Preclinical Medicine, University of Regensburg, Regensburg, Germany, [2]Centre for Electron Microscopy, Faculty for Biology and Preclinical Medicine, University of Regensburg, Regensburg, Germany

The regulation of archaellation, the formation of archaeal-specific cell appendages called archaella, is crucial for the motility, adhesion, and survival of archaeal organisms. Although the heavily archaellated and highly motile *Pyrococcus furiosus* is a key model organism for understanding the production and function of archaella in Euryarchaea, the transcriptional regulation of archaellum assembly is so far unknown. Here we show that the transcription factor EarA is the master regulator of the archaellum (*arl*) operon transcription, which is further modulated by intergenic transcription termination signals. *EarA* deletion or overexpression strains demonstrate that EarA is essential for archaellation in *P. furiosus* and governs the degree of archaellation. Providing a single-molecule update on the transcriptional landscape of the *arl* operon in *P. furiosus*, we identify sequence motifs for EarA binding upstream of the *arl* operon and intergenic terminator sequences as critical elements for fine-tuning the expression of the multicistronic *arl* cluster. Furthermore, transcriptome re-analysis across different *Thermococcales* species demonstrated a heterogeneous production of major archaellins, suggesting a more diverse composition of archaella than previously recognized. Overall, our study provides novel insights into the transcriptional regulation of archaellation and highlights the essential role of EarA in *Pyrococcus furiosus*. These findings advance our understanding of the mechanisms governing archaellation and have implications for the functional diversity of archaella.

KEYWORDS

Archaea, archaellum, transcriptomics, EarA, *Thermococcales*, single-molecule sequencing, transcriptional regulator

1. Introduction

Extracellular locomotive structures are ubiquitous across all domains of life and known as cilia or flagella in Eukaryotes, flagella in Bacteria, and archaella in Archaea (Khan and Scholey, 2018; Beeby et al., 2020). Though analogous in function to flagella, the archaellum is distinct in its evolutionary origin, genetic makeup, and the structural organization of its motility machinery (Jarrell et al., 2021; Nuno De Sousa Machado et al., 2022). Sharing similarities to the bacterial Type IV pili system, the archaellum distinguishes itself by its unique assembly and ATP-dependent rotation mechanism, in stark contrast to bacterial flagella, which operate based on ion gradients (Albers and Jarrell, 2015; Poweleit et al., 2016). Notably, in some archaea, the

function of archaella extends beyond swimming and these cell appendages contribute to surface adhesion and biofilm formation through intra- and interspecies cell–cell contacts. For instance, in the hyperthermophilic model archaeon *Pyrococcus furiosus*, archaella have been demonstrated to facilitate the attachment of cells to black smoker material taken from their native habitat, indicating their role in colonizing extreme environments (Näther et al., 2006; Schopf et al., 2008; Wirth, 2017; Wirth et al., 2018).

The complex architecture of the archaellum encompasses multiple components, including filament proteins (ArlA and/or ArlB), stator elements (ArlF, ArlG), proteins potentially needed for motor function (ArlC, ArlD, ArlE, ArlX), ATPase and ATPase-modulating proteins (ArlI, ArlH), and platform proteins (ArlJ) [for an overview see (Jarrell et al., 2021)]. Most of these components are encoded within the archaellum (*arl*) gene cluster. Arl gene clusters can be divided into the highly conserved Arl1 or Arl2 cluster that differ in their composition and phylogeny (Desmond et al., 2007). However, genes encoding associated proteins, such as enzymes for N-glycosylation of archaellins and chemotaxis system components, may be located in different regions of the genome (Schlesner et al., 2009; Jarrell et al., 2014).

To fine tune the expression of *arl* genes during different environmental conditions, diverse transcriptional regulation mechanisms have been discovered in Crenarchaeota (also known as Thermoproteota) (Jarrell et al., 2021; Oren and Garrity, 2021). In *Sulfolobus*, transcription factor (TF)-based regulation has been extensively studied, revealing several factors that control archaellum formation. Transcription factors ArnA and ArnB play key roles in this process (Reimann et al., 2012). Other regulators that either impair (ArnC, ArnS, Abfr1, ArnR, ArnR1) or enhance (ArnD, PP2A) motility have also been identified (Lassak et al., 2013; Haurat et al., 2017; Hoffmann et al., 2017; Li et al., 2017). Notably, these proteins are absent in Euryarchaeota (also known as Methanobacteriota), where only a single regulator, the euryarchaeal archaellum regulator (EarA), was recently identified (Ding et al., 2016, 2017b).

This factor is widely distributed in the Euryarchaeota and has been shown to be indispensable for transcription of the *arl* operon in *Methanococcus maripaludis*, which encodes most proteins involved in archaellum formation (Ding et al., 2016, 2017b). Interestingly, spontaneous mutation in the promoter region of an EarA knockout strain restored archaellum functionality (Ding et al., 2017a). Although the molecular mechanism was not revealed yet, it led to the hypothesis that EarA might aid in recruiting transcription factor B to a weak B recognition element (BRE) in the wild-type *arl* promoter, but may not be essential when the *arl* promoter has a strong BRE (Ding et al., 2017a). While a deletion mutant could be restored when EarA homologs from other archaea were heterologously expressed in *M. maripaludis*, this was not possible with the EarA variant from *Pyrococcus furiosus*, which could be attributed to lower sequence identity or codon usage issues (Ding et al., 2017b). This is particularly intriguing because among Euryarchaeota, *Pyrococcus* has served as a key model system to further our understanding of the archaellum machinery at the functional, structural, and regulatory levels (Näther et al., 2006; Näther-Schindler et al., 2014; Lewis et al., 2015; Daum et al., 2017; Chaudhury et al., 2018).

Given this discrepancy, our study investigated the role of the euryarchaeal archaellum regulator EarA in *P. furiosus*, finding that EarA is also essential for archaellation in this hyperthermophilic organism. Additionally, we present an updated single-molecule perspective of the transcriptional landscape of the *arl* operon. Finally, we contextualized our findings by integrating transcriptomics data from across the *Thermococcales* and examining the potential influence of sequence elements on the expression of the *arl* operon. This includes the presence of EarA binding sequences upstream of the *arl* promoter, as well as intergenic terminator sequences. Together these data suggest that while the transcription factor EarA is the key regulatory player of archaellum expression, intergenic sequences help in fine-tuning other components of the *arl* operon in *Thermococcales*.

2. Materials and methods

2.1. Strains, plasmids, primers, and antibodies

All strains, plasmids, primers, and antibodies used in this study are listed in Supplementary Table 1.

2.2. Construction of vectors and transformation of *Pyrococcus furiosus*

For the construction of *P. furiosus* strains Δ*earA*, Δ*earA* recovery and Δ*earA* overexpression, the modified genetic system for *P. furiosus* DSM3638 based on selection via agmatine auxotrophy as described in Grünberger et al. (2021) was used.

For markerless disruption of the *P. furiosus earA* gene, the upstream and downstream flanking regions of the *P. furiosus earA* gene including the promoter region were amplified using the primer pairs 0340upAscIFW/0340upFusRW and 0340doFW/0340doNotIRW (Supplementary Table 1). Both PCR products were used as a template for a single-overlap extension PCR and the resulting PCR product was cloned into the modified pMUR47 vector containing a two-gene resistance cassette via the AscI and NotI restriction sites (Grünberger et al., 2021). Successful and correct insertion of the PCR fragment into the vector was verified by DNA sequencing resulting in the plasmid pMUR443.

Circular plasmid DNA pMUR443 and strain MURPf37 were used for transformation and selection was carried out in 1/2 SME-starch liquid medium without agmatine sulfate and inosine+guanosine (I+G) at 85°C for 12 h (Grünberger et al., 2021). Pure cultures of the intermediate mutant MURPf76_i were obtained by plating the cells on solidified medium. The integration of the plasmid into the genome by single cross-over was verified by analyzing corresponding PCR products. For the counter selection cells were plated on solidified medium containing 50 µM 6-methylpurine and 8 mM agmatine sulfate to induce a second homologous recombination step to recycle the selection marker and to eliminate integrated plasmid sequences. The genotype of the final mutant (MURPf76) was confirmed by PCR and cells had to be grown in the presence of 8 mM agmatine sulfate and 8 mM I+G.

For EarA recovery and EarA overexpression in trans, a shuttle vector system for *P. furiosus* was used. Based on plasmid pMUR310 (Grünberger et al., 2021), an overexpression shuttle vector system was constructed that allows constitutive expression of a protein of interest via the *gdh* promoter (PCR amplified from *T. kodakarensis* genomic DNA using primers RPA_pYS13_GA_F/pYS13_GA-R) and the *hypA1* gene terminator (PCR amplified from *P. furiosus* genomic DNA using primers PYS14_GA_F/pYS13_RPA_GA_R) regions. Both regions were fused upstream and downstream of the gene of interest as described in

Waege et al. (2010) and cloned into pMUR310 via the EcoRV restriction site. One of these overexpression plasmids was used for PCR amplification of the plasmid backbone using primers pYS14Expgdh_F/ pYS14_GA_F. The *earA* gene was PCR amplified from *P. furiosus* genomic DNA using primers flaaktivGAp914F/ flaaktivgap914r. Both PCR products were assembled using Gibson Assembly (Gibson et al., 2009) resulting in the plasmid pMUR444. Next, the Strep-tag® II sequence was inserted at the C-terminus via Phusion® High-Fidelity DNA Polymerase mutagenesis PCR (New England Biolabs) using the primers pYS14_GA_F/ 5′ phosphorylated flaakstrp14R to amplify the whole plasmid. The PCR product was ligated using T4 ligase (New England Biolabs) and transformed into *E. coli* DH5α cells (New England Biolabs), resulting in pMUR445. The final construct including the correct insertion of the Strep-tag® II sequence was verified by sequencing (Microsynth). Finally, to construct a shuttle vector expressing EarA under its native promoter the pMUR444 backbone including the *earA* gene, but without *gdh* promoter, was PCR amplified using primers flaaktivpromGAp914F/GAp14bbminpromR. The *earA* promoter region was PCR amplified from *P. furiosus* genomic DNA using primers p14fusPF0340prom/034upRW1. Both PCR products were assembled using Gibson Assembly (Gibson et al., 2009) resulting in the plasmid pMUR458. The correct exchange of promoter regions was verified by sequencing (Microsynth).

1 µg of the circular plasmids pMUR310, pMUR445 and pMUR458 were transformed into MURPf76 as described previously (Waege et al., 2010; Kreuzer et al., 2013; Grünberger et al., 2021). Selection was carried out in 1/2 SME liquid medium without agmatine sulfate and I + G at 85°C for 12 h. Pure cultures of the mutants MURPf79 (Δ*earA* strain containing pMUR310), MURPf77 (Δ*earA* recovery strain containing pMUR458) and MURPf78 (Δ*earA* overexpression strain containing pMUR445) were obtained by plating the cells on solidified medium. Plasmid stability was verified by re-transformation into *E. coli* DH5α and DNA sequencing of purified plasmids (Microsynth). Final *P. furiosus* mutants could be grown in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract but without agmatine sulfate and I + G supplementation (Grünberger et al., 2021).

2.3. Analysis of swimming behavior

Pyrococcus furiosus strains were grown at 85°C in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract to a cell density of ~1×10^6 cells per mL and swimming experiments were performed using a temperature gradient forming device (TGFD) at 100°C as described in Mora et al. (2014). Movies were recorded and analyzed as described previously (Herzog and Wirth, 2012; Mora et al., 2014).

2.4. Western blot analysis

Pyrococcus furiosus strains were grown at 85°C in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract to a cell density of ~1×10^8 cells per mL and harvested by centrifugation at 20,500 *g* (4°C). The cell pellets from 20 mL cultures were resuspended in 1x PBS buffer and the cell suspensions were sonicated for 50 s using a Bandelin Electronic™ Sonopuls™ HD 3400 homogenizer (cycle: 80%, power 80%). After centrifugation for 30 min at 16,200 *g* (4°C), total protein concentrations of the cell extracts were quantified using the QubitTM Protein Assay Kit 3.0 Fluorometer (ThermoFisher). 2 µg of total protein extracts from each *P. furiosus* strain were separated using SDS-PAGE. Western blots were performed as described previously (Waege et al., 2010).

2.5. PCR test

Pyrococcus furiosus strains were grown at 85°C in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract to a cell density of ~1×10^8 cells per mL and harvested by centrifugation at 20,500 *g* (4°C). Genomic DNAs were isolated using the ReliaPrep™ gDNA Tissue Miniprep System (Promega) following the Mouse tail protocol. The region of the *arlD* gene was PCR amplified using the primers Fladup_100fw/ Fladdo_100rw, and the region of the *earA* gene locus was PCR amplified using the primers 0340do_113rw/0340up_100fw, respectively. PCR products were separated by agarose gel electrophoresis.

2.6. Transmission electron microscopy

Pyrococcus furiosus cells grown to the exponential phase were taken up in a syringe and carefully filtered through a polyamide/nylon filter (pore size 0.2 µm). The concentrated cell film was resuspended in 20–40 µL fresh media and chemically fixed with 1% (v/v) glutaraldehyde (final concentration) at room temperature. 5 µL of the fixed cell suspension were applied onto copper grids (400-mesh; G2400C, Plano) coated with a 10 nm carbon film (Carbon Coater 208 Turbo, Cressington), washed twice by blotting with distilled water, and the samples were subsequently shadowed with ~1.0 nm Pt/C (15° angle; CFE-50, Cressington). Transmission electron micrographs were recorded on a CM12 transmission electron microscope (FEI) operated at 120 keV and fitted with a slow-scan CCD camera (TEM 0124; TVIPS) using EMMENU v4.0 (TVIPS).

2.7. Gene calling and re-annotation

All currently available *Thermococcales* RefSeq reference genomes with a completeness of scaffold or better were acquired via the NCBI genome page (compare Supplementary Table 2).[1]

The genomes were annotated via the arCOG definitions as described elsewhere (Dombrowski et al., 2020). In short, CDS sites and protein sequences predicted with Prokka (v. 1.14.6) (Seemann, 2014) were searched against arCOG[2] (Makarova et al., 2015) Hidden Markov Models (HMM) prepared with hmmbuild using hmmsearch (HMMER v. 3.3.2; Nov 2020).[3] The best hit was assigned as annotation based on bit score then e-value, after an e-value cutoff <=1e-3 (see Supplementary Table 2).

1 https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=2258&reference_only=true&assembly_level=1%3A3; accessed 2023-05-12.

2 https://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG/zip.aliar14.tgz

3 http://hmmer.org/

2.8. Re-analysis of short-read RNA-seq data

For coverage analysis of the *arl* gene cluster the following RNA-seq datasets were reanalyzed: lllumina data from *Pyrococcus furiosus* DSM3638 exponential growth phase at the optimum growth temperature of 95°C (ERR11200497, ERR11200498, ERR11200499, ERR11200500) (Grünberger et al., 2023), *Palaeococcus pacificus* DY20341 cells from exponential growth phase with S^0 addition (SRR10749090) (Zeng et al., 2020), *Thermococcus onnurineus* NA1 cells from exponential phase in CO-supplied media (SRR1702360, SRR1702361) (Lee et al., 2015), *Thermococcus litoralis* DSM5473 cells from exponential phase (SRR12486532, SRR12486546, SRR12486547) (Liang et al., 2021) and *Thermococcus kodakarensis* KOD1 cells from exponential phase (ERR6384078, ERR6384079, ERR6384080, ERR6384081) (Villain et al., 2021).

Sequencing reads in fastq format were filtered and trimmed using fastp (v. 0.23.3) to remove low-quality reads and adapter sequences (−-cut_front, −-cut_tail, −-qualified_quality_phred 30, −-length_required 30) using parameters selective for single-end or paired-end reads (Chen et al., 2018). Subsequently, trimmed reads were aligned to the respective NCBI reference genomes (*P. furiosus*: NZ_CP023154.1, *P. pacificus*: NZ_CP006019.1, *T. onnurineus*: NC_011529.1, *T. litoralis*: NC_022084.1, *T. kodakarensis*: NC_006624.1) using Bowtie2 (v.2.5.1) with default parameters (Langmead and Salzberg, 2012). The resulting sequence alignment files (SAM) were converted to binary mapping format (BAM) and sorted using samtools (v.1.17) (Li et al., 2009). Additionally, position-specific coverage files were generated using samtools depth (−a −J), including reads with deletions in the coverage computation. Downstream normalization and visualization was performed using the Tidyverse in R (Wickham et al., 2019). Briefly, coverage for each position was first normalized by counts per million (CPM), before calculating the mean value of all replicates.

For calculation of transcripts per million (TPM) values of the protein-coding genes, featureCounts (RSubread package v. 2.10.5) was used to calculate the count matrix based on custom GFF files containing all *arl* genes detected using HMM as described above (Liao et al., 2019). Next, TPM values were calculated by dividing the number of reads mapping to each gene by the gene length in kilobases, then dividing the resulting reads per kilobase (RPK) values by the sum of all RPK values in the sample, and finally multiplying the quotient by one million. For visualization, the z-score was calculated for all genes in the *arl* gene cluster region, including the two upstream genes.

2.9. Re-analysis of Nanopore RNA-seq data

Nanopore data from two biological replicates of *Pyrococcus furiosus* DSM3638 cells from exponential growth phase grown at the optimum growth temperature of 95°C were reanalyzed (ENA-project: PRJEB61177, runs: ERR11203080, ERR11203081) (Grünberger et al., 2023). Generation of the dataset, including detailed information about cell growth, RNA isolation, RNA treatment, library preparation, sequencing following the PCR-cDNA barcoding kit protocol (SQK-PCB109) from Oxford Nanopore Technologies (ONT) and data analysis is described in Grünberger et al. (2023). Briefly, basecalling and demultiplexing was performed using guppy (v. 6.4.2 + 97a7f06) in high-accuracy mode. Full-length sequenced reads were detected, strand-oriented and trimmed using pychopper (v.2.7.2)[4] with autotuned cutoffs and the edlib backend for identifying the custom VNprimer(5′-ACTTGCCTGTCGCTCTATCTTCATTGATGGTGCC TACAG-3′). Mapping to the *P. furiosus* DSM 3638 genome (NCBI:NZ_CP023154.1) was performed using minimap2 (v. 2.24-r1122) with standard parameters suggested for aligning Nanopore genomic reads (−ax map-ont) (Li, 2018, 2021). Soft or hard clipped (>5 bases) reads were removed using samclip (v. 0.4.0), and SAM files were converted to sorted BAM files using samtools (v. 1.16.1) (Li et al., 2009). Coverage files generation including samtools depth and CPM normalization in R was performed as described for the short-read datasets. Single-read tracks of unspliced reads (njunc = 0) were plotted using ggplot2 geom_segment from read start to end.

2.10. Sequence analysis and motif identification

Annotations were filtered using R and the Tidyverse packages (v. 2.0.0) (Wickham et al., 2019). Genomic sequences annotated as EarA were extracted and translated using seqkit (v. 2.4.0) (Shen et al., 2016). EarA protein sequences were aligned using ClustalO (v. 1.2.3) (Sievers and Higgins, 2014) with default parameters and visualized using ggmsa (v. 1.4.0) (Zhou et al., 2022).

The genomic locations that matched the "FlaB" HMM were annotated as "arlB" and assigned a rank, with the most-upstream location designated as "arlB0" to be consistent with the annotation used in *Pyrococcus furiosus*. Please note that this deviates from the database annotation of other strains that start with "arlB1." Genomic regions located 150 nt upstream of *arlB*0 and 50 nt downstream of all *arlB* genes were defined using plyranges (v. 1.18.0) (Lee et al., 2019). The sequences corresponding to these regions were extracted using seqkit subseq.

Motif analysis was conducted on the sequences upstream of the *arlB*0 genes using MEME (v. 5.5.2; −mod anr −dna −w 8) (Bailey et al., 2015). The identified motifs and their respective locations were visualized using the ggmotif (v. 0.2.1) (Li et al., 2022), ggseqlogo (v. 0.1) (Wagih, 2017), memes (v. 1.6.0), and Tidyverse packages.

The enrichment of "T" bases in the genomic regions downstream of the *arlB* genes was calculated as the $\log_2$ fold change of the letter frequency in the 50 nt downstream region, obtained using bedtools nuc (v. 2.31.0) (Quinlan and Hall, 2010), to the average letter frequency in the whole genome retrieved using the 'letterFrequency ("T")' function from the Biostrings package (v. 2.66.0).

3. Results

3.1. The euryarchaeal archaellum regulator EarA is essential for archaellation in *Pyrococcus furiosus*

To elucidate the *in vivo* role of the regulator EarA, we took advantage of the existing genetic system of *P. furiosus* based on

4 https://github.com/epi2me-labs/pychopper

agmatine auxotrophy (parental wildtype) (Waege et al., 2010) and constructed a mutant strain (Δ*earA*) in which the *earA* gene locus including the corresponding promoter region was deleted. Gene rescue strains expressed EarA *in trans* either under the control of its native promoter (Δ*earA* recovery) or under the control of the strong, constitutive *gdh* promoter (Δ*earA* overexpression) (Micorescu et al., 2008). Successful deletion of the genomic *earA* locus was confirmed via PCR (Supplementary Figure 1A). Additionally, we assessed the recovery and overexpression of EarA levels within the Δ*earA* background by quantifying the production of archaellins using Western blot analysis (Supplementary Figures 1B,C).

The archaellation of these strains was visualized by transmission electron microscopy of Platinum/Carbon shadowed samples (Figures 1A–D; Supplementary Figure 2). The *earA* gene deletion led primarily to cells devoid of archaella, whereas overexpression resulted in heavily archaellated *P. furiosus* cells (Figures 1A–D). These outcomes mirrored the observed swimming behaviors, which indicated that no or few archaella in the deletion mutant are not sufficient to allow the cells to swim normally (data not shown).

To systematically evaluate the observed initial differences in the extent of archaellation, we pursued further analysis by classifying the archaellation level based on the number of visible archaella of approximately 70–90 cells for each strain used in this study as none (0), low (1–10), mid (11–40) and high (>40) (Figure 1E; Supplementary Figure 2A). The increased EarA protein levels observed in the Δ*earA* overexpression strain indeed correlated with a significantly higher *in vivo* archaellation level clearly surpassing both the parental wildtype and the recovery strain (Figure 1F; Supplementary Figure 2B). These results underscore EarA as a pivotal regulator of archaellation in *P. furiosus*, suggesting that the cellular levels of this transcriptional regulator directly govern the degree of archaellation.

3.2. A single-molecule update on the transcriptional regulation of the *arl* gene cluster in *Pyrococcus furiosus*

To explore the potential hierarchic nature of EarA-induced transcriptional regulation, we analyzed the transcriptome landscape of the *arl* gene cluster in *Pyrococcus furiosus*. Therefore, we re-analyzed both short- and long-read RNA-seq data in addition to other database information, including transcription start and termination sites (Figure 2; Grünberger et al., 2019, 2023). The Arl1-type cluster is characterized by three archaellin genes (*arlB0*, *arlB1*, *arlB2*) followed by the *arl* accessory genes *arlC*, *arlD*, *arlF*, *arlG*, *arlH*, *arlI*, *arlJ* and a hypothetical gene (Näther-Schindler et al., 2014; Jarrell et al., 2021). According to the DOOR2 database annotation, two genes preceding the *arl* genes, namely the transcriptional regulator EarA and the postulated methyltransferase *fam*, are part of the operon (Figures 2A,B). In contrast, prediction based on short-read RNA-seq data from mixed conditions suggests that the two genes are transcribed as an extra single unit (Figure 2A; Mao et al., 2014; Grünberger et al., 2019).

This was further supported by coverage re-analysis of short-read RNA-seq recently obtained from cells grown to exponential phase and cells recovered at the optimal temperature of 95°C following an extended cold shock at 4°C (Figures 2C,D; Grünberger et al., 2023). Cells recovered from such conditions have been demonstrated to respond swiftly by swimming, which was confirmed by the induction of the *arl* gene cluster (Mora et al., 2014). Interestingly, we noted a considerable upregulation of the EarA regulator ($\log_2$ fold change: 2.3), initiating the activation of the *arl* genes. The level of activation decreased as the distance from the primary transcription start site increased.

To achieve a higher resolution of the transcriptional output, we re-analyzed long-read single-molecule data from samples cultivated under optimal growth conditions (Grünberger et al., 2023). Previous transcriptional start site mapping using differential RNA sequencing of Terminator-Exonuclease (TEX) treated RNA revealed that the primary start site is situated upstream of *arlB0*, the essential major archaellin protein (Grünberger et al., 2019). Indeed, the majority of the reads from long-read sequencing originate from this position. Furthermore, long-read coverage suggests differential expression of multiple transcription units, prompting us to refine and update the transcription model based on RT-PCR analysis (Figures 2E–H; Näther-Schindler et al., 2014). In alignment with a previous study and structural data, our analysis verified that *arlB0* is the primary archaellin transcript, particularly during early exponential growth (Näther-Schindler et al., 2014; Daum et al., 2017). However, *arlB0* can be detected as co-transcript with downstream genes but also as a single transcript. Notably, we discovered several shortened units derived from the primary *arl* operon transcript (Figure 2H). These units exhibited a subset of reads originating from the upstream position of *arlB0*. Moreover, we detected transcripts with random downstream read starts that could not be attributed to promoter elements, suggesting the occurrence of 5′ processing events. Our findings revealed that transcripts spanning from *arlB0* to *arlJ* were not detected, in agreement with the previous study (Näther-Schindler et al., 2014). Importantly, the absence of enriched 5′ ends in our dataset, including the TEX-enriched deep-sequenced RNA-seq data from cells under various conditions, strongly suggests that all transcripts likely originate from *arlB0*. The hierarchical gene expression changes observed in the cold-shocked recovery cells further support this conclusion.

To validate these findings, we performed additional analysis on enriched 3′ ends and conducted a terminator sequence analysis (Figures 2I–K). This analysis confirmed the observed coverage profiles and highlighted the presence of poly(T)-sequences after *arlB0*, *arlB2*, and *arlD* that may serve as intergenic transcriptional terminators.

3.3. Investigation of transcriptome data and sequence features in *Thermococcales* reveal a heterogeneous production of major archaellins

We expanded our study by reanalyzing RNA-seq data from various organisms within the order *Thermococcales* including *Palaeococcus pacificus*, *Thermococcus onnurineus*, *Thermococcus kodakarensis*, and *Thermococcus litoralis* to understand the transcriptional characteristics of the *arl* gene cluster (Lee et al., 2015; Zeng et al., 2020; Liang et al., 2021; Villain et al., 2021; Grünberger et al., 2023). While the overall Arl1-like operon organization is well conserved among these organisms, they exhibited variations in the number of *arlB* genes, ranging from two

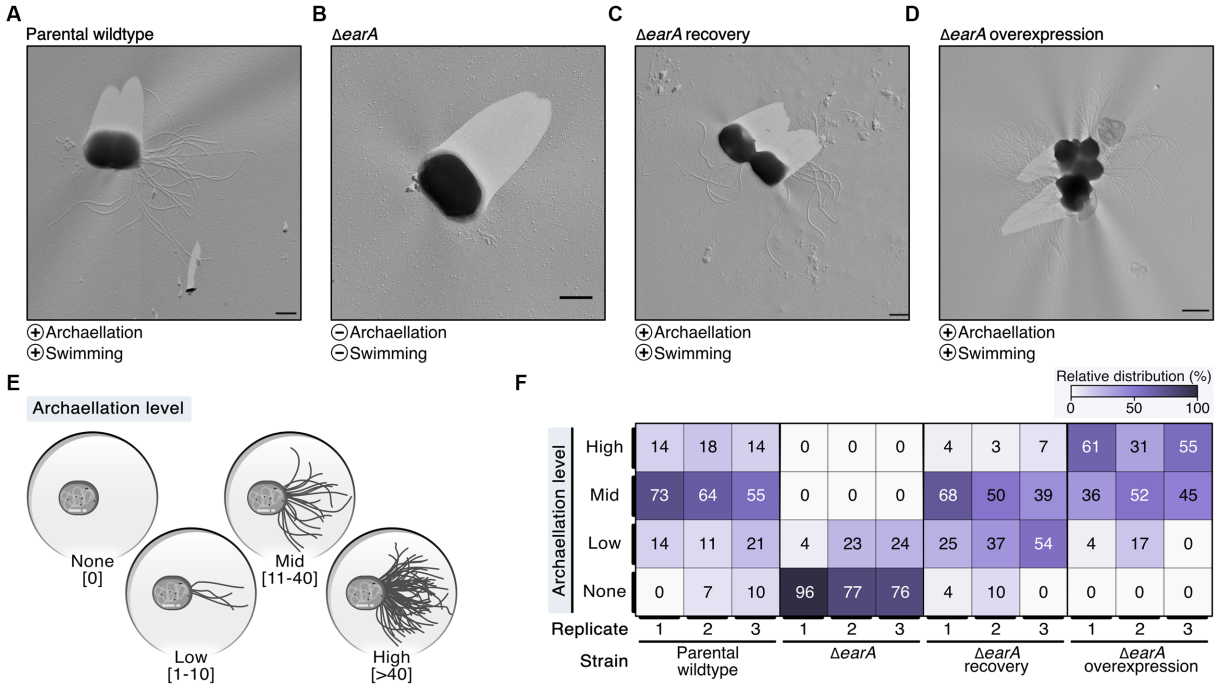

FIGURE 1
Electron microscopy evaluation of archaella formation in different genetic variants of *Pyrococcus furiosus* illustrates the central role of EarA. **(A)** Transmission electron microscopy (TEM) micrographs of representative cells observed in the parental wildtype, **(B)** Δ*earA* knockout strain, **(C)** Δ*earA* recovery (native promoter) and **(D)** Δ*earA* overexpression strain (*gdh* promoter). The presence and absence of archaellation and swimming behavior are denoted by plus and minus symbols, respectively. Scale bar: 0.5 μm. **(E)** Systematic classification of the archaellation level based on the number of observable archaella originating from one cell. **(F)** Color-coded representation (from 0% in white to 100% in violet) of the relative distribution of the four archaellation levels across the strains under investigation.

to five gene variants (Figure 3A). Broadening the genomic analysis, in all but six of the 33 *Thermococcales* reference genomes currently available in the NCBI genomes page, arCOG Hidden Markov Model (HMM) matches for the genes encoding for EarA and Arl proteins were found (Supplementary Figure 3; Supplementary Table 2). The transcriptional regulator EarA is highly conserved in *Thermococcales* and mostly oriented in the same orientation as the arlB cluster, with some exceptions (Supplementary Figures 3, 4; Ding et al., 2017b). Note, that the NCBI annotations for the six reference genomes without HMM matches do not contain archaellin annotations and three of these organisms were described as "non-motile" or "not-flagellated" (Canganella et al., 1998; Grote et al., 1999; Hensley et al., 2014).

Comparing the coverage profiles of *Pyrococcus* with the other strains, we consistently observed a higher number of transcripts in the *arlB* region compared to the downstream accessory *arl* genes (Figures 3A,B). This finding indicates that under exponential growth conditions, *arlB* transcripts are the most abundant in all *Thermococcales*. In contrast to *Pyrococcus*, there were notable differences in the transcriptional output of the other strains. Notably, *arlB0* was not consistently the major archaellin based on expression levels. We observed some variability of the transcript levels, with either *arlB1* or *arlB2* as the most highly expressed gene or a balanced distribution between the *arlB* genes (Figures 2A,B). This balance was particularly evident for the five tandem co-transcribed archaellin genes in *T. kodakarensis*, aligning with previous Northern Blot results (Nagahisa et al., 1999). Also, the transcript counts of the regulator *earA* were significantly higher or equal to the *arlB* expression levels in the *Thermococcus* strains, suggesting it is transcribed as a separate unit.

Since previous analysis showed that EarA binds to a 6 bp consensus sequence (TACATA) in the promoter region of the first archaellin gene in *Methanococcus*, we performed sequence analysis in all *Thermococcales* (Ding et al., 2017b). We found that the *arl* loci in all organisms examined were preceded by archaea-typical BRE and TATA elements, which were uniformly located at a similar distance from the start codon of the first *arl* gene (Figures 3C,D). Specifically, the *arlB0* gene in *P. furiosus* featured a 30 nt long 5' UTR (Grünberger et al., 2019).

Moreover, most cases showed a motif A(C/G)(C/G)TACA(T/C) at a fixed distance of nine nucleotides upstream of the BRE element, often repeating up to three times adjacently. We observed three distinct patterns, with either one (e.g., *P. horikoshii*), two (*P. furiosus*), or three EarA recognition sites (*T. kodakarensis*). It is noteworthy that the motifs in all *Pyrococcus* strains were shifted to either one window upstream or downstream compared to those in most of the *Thermococcus* strains. The absence of this motif in the upstream regions of all other *arl* genes supports the notion that EarA specifically binds upstream of the *arlB0* gene, thereby regulating the full gene cluster.

As a second layer of transcriptional regulation, we observed internal poly(T) sequences in *P. furiosus* (Figures 2I–K). Investigating whether this also applies to other *Thermococcales* and whether the strongest potential terminator sequence is positioned after the major archaellin, we calculated the enrichment of poly(T) sequences located

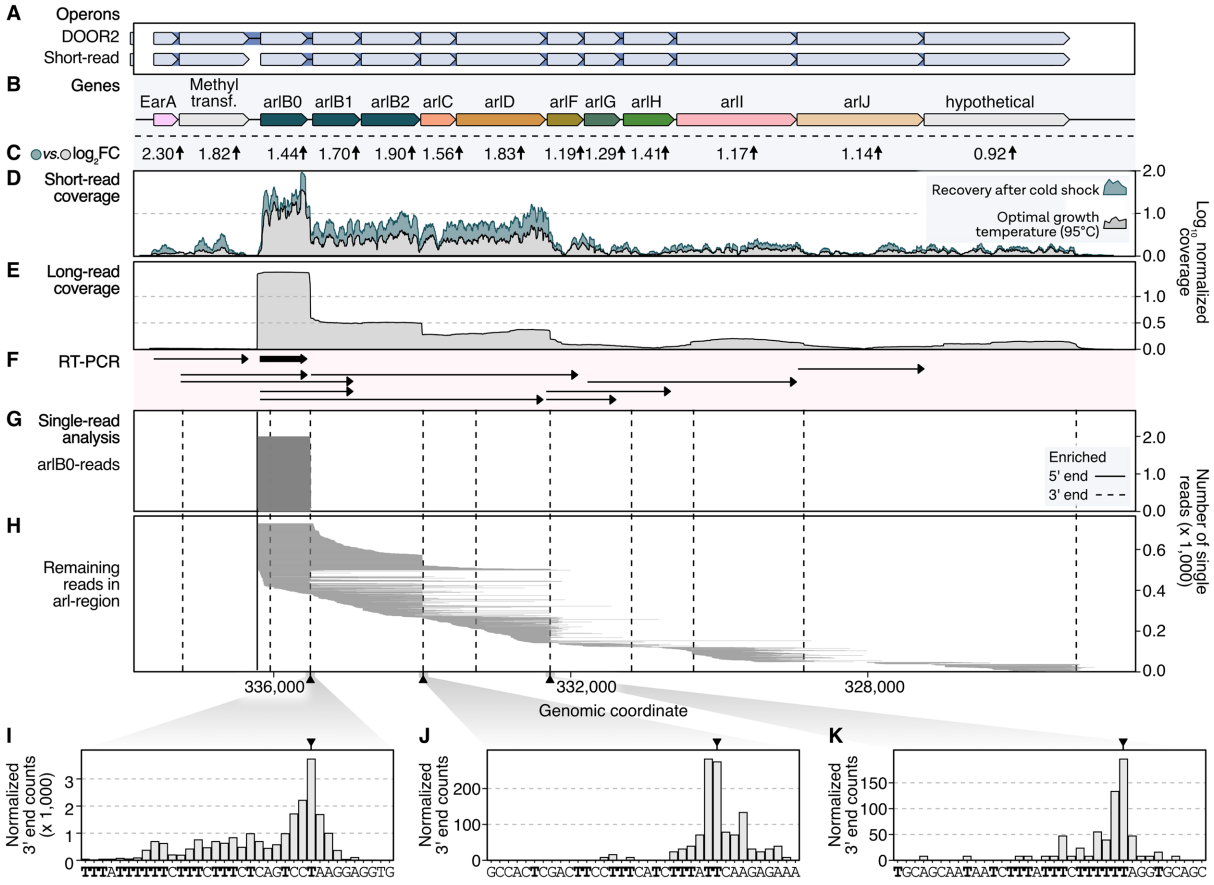

FIGURE 2
Long-read sequencing analysis provides high-resolution insights into transcriptional regulation of the *arl* gene cluster in *P. furiosus*. **(A)** Annotated operons in the DOOR2 database and as suggested by ANNOgesic using short-read RNA sequencing of mixed growth conditions (Mao et al., 2014; Grünberger et al., 2019). **(B)** Gene annotation of the *arl* cluster, including the transcriptional regulator EarA and a postulated methyltransferase upstream of the *arlB0* gene. **(C)** Log_2 fold changes obtained from differential gene expression analysis of samples recovered at 95°C after prolonged cold shock at 4°C (24h) compared to cells in exponential growth phase (Grünberger et al., 2023). **(D)** Mean coverage profiles normalized by CPM per position of these conditions are shown in gray (exponential phase) and blue (recovery) (Grünberger et al., 2023). **(E)** Mean long-read sequencing coverage profile (Nanopore SQK-PCB109) of cells in exponential growth phase. **(F)** RT-PCR results from Näther-Schindler et al. (2014). Detected transcripts are visualized as arrows. *ArlB0* was detected as the major archaellin in *P. furiosus* and is highlighted in bold. **(G)** Single-read plotting of reads mapping to *arl* region, split by reads mapping to *arlB0* and **(H)** all other reads. Enriched primary 5′ ends from short-read sequencing of TEX-treated RNAs (Grünberger et al., 2019) and primary 3′ ends obtained from a recent Term-seq dataset (Grünberger et al., 2023) are shown as solid and dashed vertical lines, respectively. **(I)** Histograms of CPM-normalized 3′ end positions in three regions [downstream of *arlB0*: **(I)** downstream of *arlB2*: **(J)** downstream of *arlD*: **(K)**] that showed significant enrichment of 3′ ends using short-read Term-seq downstream of coding genes (Grünberger et al., 2023).

50 bases downstream of each *arlB* gene. Our results aligned with the overall expression levels, showing the highest T enrichment after the most abundant *arlB* gene (Figure 3E). The analysis across all *Thermococcales* confirmed our coverage analysis and indicated that, despite the *arl* gene organization, the most upstream *arl* gene is not always the most abundant gene, supporting the hypothesis that different ArlB might get incorporated into final archaellum filaments.

4. Discussion

In this study, we highlighted the crucial role of the euryarchaeal archaellum regulator EarA for archaellation within *P. furiosus*. Using TEM analysis of different genetic variants, we gathered *in vivo* evidence that the expression levels of EarA directly correlate with the level of archaellation. Given that the production of each protein component of the archaellum needs to be carefully controlled, we investigated how this control might be achieved on the transcriptional level (Chaban et al., 2007).

Our findings provide experimental support for prior speculations about EarA-based activation of transcription. Based on the occurrence of a spontaneous promoter mutation it was postulated that EarA might aid in the recruitment of TFB to a weak BRE element, marked by 7–8 consecutive adenines, in the *arl* gene cluster's promoter (Ding et al., 2017a). This hypothesis is supported by our motif analysis which reveals a similarly weak BRE sequence present in all *Thermococcales* preceding the first *arl* gene. Notably, the approximated spacing of 9–11 base pairs between the first upstream located EarA binding site and the BRE implies a direct interaction of EarA with TFB. This interaction is reminiscent of the activation pattern observed for the potent transcriptional activator 2 (Ptr2), which is presumed to interact directly with TBP to recruit it

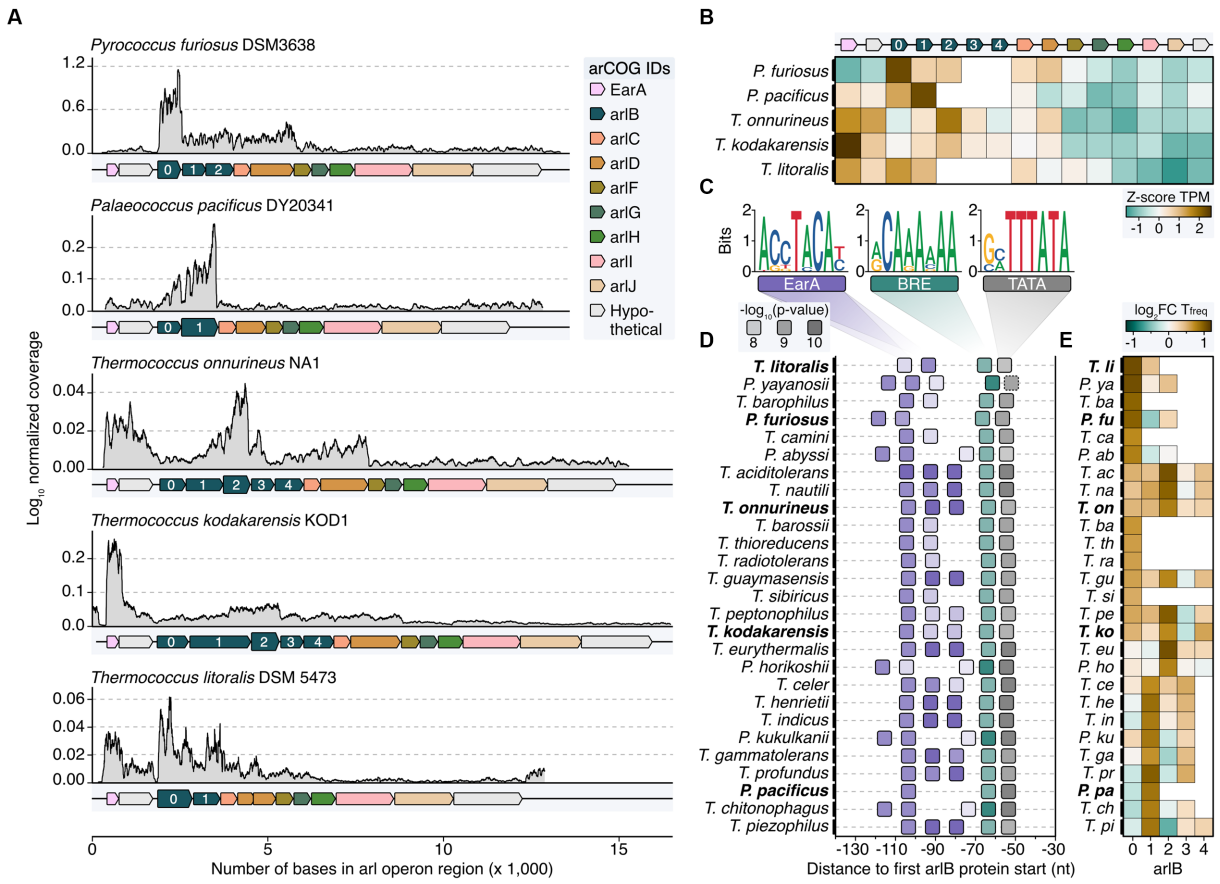

FIGURE 3
Transcriptome and sequence feature analysis reveal heterogeneous expression profiles of major archaellins in *Thermococcales*. **(A)** Normalized coverage profiles from short-read sequencing of archaeal cells grown until exponential growth phase. Note that throughout the manuscript the most upstream *arl* gene was renamed *arlB0* in line with the current annotation in *P. furiosus*. Highest expressed *arlB* transcripts are highlighted in bold. **(B)** Z-score normalized TPM values (brown: high TPM, green: low TPM) for all genes shown in panel **(A)**. **(C)** Enriched motifs potentially recognized by EarA (pink), TFB (BRE, green) and TBP (TATA, gray) identified using MEME search (Bailey et al., 2009) of the sequence region 150 nt upstream of the first *arl* gene. **(D)** Position of detected motifs (colored according to the description in **C**) are shown in a window from position −30 to −140 from the first *arlB* gene start. *P*-values are indicated by transparency of the boxes. Organisms shown in panels **(A,B)** are highlighted with bold text. The dashed box outline of the TATA box indicates that the element was not found by MEME but added after manual inspection. **(E)** Enrichment of the base "T" 50 nt downstream of the respective *arlB* gene over genomic "T" frequency. Boxes are filled according to the $\log_2$ fold enrichment in brown (high "T"-frequency) or green (low "T"-frequency).

to an imperfect TATA-box (Ouhammouch et al., 2005). Until now, only one archaeal transcriptional regulator, TFB recruitment factor 1 (TFB-RF1), that mediates TFB-recruitment, has been biochemically characterized in depth (Ochs et al., 2012). For both TFB-RF1 and EarA, it remains elusive which internal and/or external cellular stimuli impact their activities as transcriptional regulators (Reichelt et al., 2018). This aspect warrants more detailed examination, especially in organisms that encode both, the archaellum machinery and chemotaxis systems, such as *T. kodakarensis* (Briegel et al., 2017). To address this, strains with targeted knockouts of EarA could be generated, followed by a comprehensive transcriptome analysis to assess global changes in gene expression and identify genes directly regulated by EarA. Additionally, a systematic approach of individually knocking out *arl* or chemotaxis genes, followed by phenotypic and transcriptomic analyses, would enable the identification of downstream effects and potentially the construction of a regulatory network, contributing to a better understanding of its cellular functions.

In contrast to transcriptional regulation at the level of initiation, relatively little is known about fine-tuning gene expression using terminator sequences, particularly in operons. In Archaea, poly(T) sequences have been shown to mediate intrinsic termination, as well as the recruitment of the recently identified termination factor aCPSF1, which is involved in 3′ end formation (Maier and Marchfelder, 2019; Sanders et al., 2020; Yue et al., 2020; Li et al., 2021; Grünberger et al., 2023). Our findings suggest that operon-internal poly(T) sequences likely function as intergenic transcriptional terminators for transcripts originating from *arlB0*. This type of terminator-based regulation and the role of readthrough in archaella accessory gene production have already been observed in *S. acidocaldarius* and appear to be important in *Thermococcales* as well (Lassak et al., 2012). Conceptually, this represents an elegant mechanism for the regulation of the multicistronic *arl* cluster allowing the fine-tuning of high-demand core components and lower-demand accessory genes (Thomas and Jarrell, 2001). However, extrapolating our findings to the final incorporation of the major

archaellin into the archaellum is not straightforward, due to the potential influences of post-transcriptional and post-translational regulation. Intriguingly, our single-molecule analysis of transcripts in *P. furiosus* showed that mRNA stability could indeed play a role, potentially explaining why the most upstream gene is not always the highest expressed.

In the case of *Thermococcus kodakarensis*, RNA-seq data compare favorably with a quantitative proteomics study. The latter disclosed similar proportions (0.4–0.25%) of total cell protein for *arlB0*, *arlB1*, and *arlB2*, hinting at a heterogeneous composition of the archaellum filament (Briegel et al., 2017). Although previous data were interpreted such that, like in Crenarchaeota, euryarchaeal archaella mainly consist of one type of archaellin only, emerging evidence contradicts this (Jarrell et al., 2021). Examples are the heterodimers of arlB1 and arlB2 observed in *Halorubrum lacusprofundi* (Pyatibratov et al., 2020) and *Methanocaldococcus villosus* (Gambelli et al., 2022). Whether this heterogeneity correlates with a structural organization of the archaellum analogous to the hook-, filament-, and cap-structures in the bacterial flagellum, or any type of multifunctionality of the archaellum, such as adhesion to cells or a variety of surfaces, warrants future investigation (Näther et al., 2006; Schopf et al., 2008; Weiner et al., 2012). Another intriguing prospect is exploring how the expression of a specific ArlB variant, presumably the principal archaellin, can be post-transcriptionally regulated if its encoding gene is not the most upstream *arlB* gene in the *arl* operon (*P. pacificus*, *T. onnurineus*).

In conclusion, our research uncovers new facets of the transcriptional regulation of the *arl* gene cluster in *P. furiosus* and establishes EarA as crucial transcriptional regulator for archaellation in *Pyrococcus furiosus*. Interestingly, the observed variation in *arlB* gene number and expression among *Thermococcales* indicates a more diverse production and assembly of archaellins than previously thought. Furthermore, our findings concerning the promoter motifs and intergenic poly(T)-sequences advance our understanding of the transcriptional regulation and potential mechanisms impacting the integration of *arl* gene products into archaellum filaments.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.ncbi.nlm.nih.gov/sra/?term=SRR10749090 (*Palaeococcus pacificus* DY20341); https://www.ncbi.nlm.nih.gov/sra/?term=SRR1304620 (*Pyrococcus abyssi*); https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200497 (*Pyrococcus furiosus*); https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200498 (*Pyrococcus furiosus*); https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200499 (*Pyrococcus furiosus*); https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200500 (*Pyrococcus furiosus*); https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384078 (*Thermococcus kodakarensis* KOD1); https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384079 (*Thermococcus kodakarensis* KOD1); https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384080 (*Thermococcus kodakarensis* KOD1); https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384081 (*Thermococcus kodakarensis* KOD1); https://www.ncbi.nlm.nih.gov/sra/?term=SRR12486532 (*Thermococcus litoralis* DSM 5473); https://www.ncbi.nlm.nih.gov/sra/?term=SRR12486547 (*Thermococcus litoralis* DSM 5473); https://www.ncbi.nlm.nih.gov/sra/?term=SRR12486546 (*Thermococcus litoralis* DSM 5473); https://www.ncbi.nlm.nih.gov/sra/?term=SRR1702360 (*Thermococcus onnurineus* NA1); https://www.ncbi.nlm.nih.gov/sra/?term=SRR1702361 (*Thermococcus onnurineus* NA1).

Author contributions

RS performed bioinformatic analysis of the data, with contributions from FG. LN performed analysis of swimming behavior and TEM imaging. RRe performed construction of vectors and genetic manipulation of *P. furiosus* cells. RRe, RRa, and DG supervised this work. DG acquired funding to conduct the work. FG prepared the figures, with contribution of all authors. RS, RRe, DG, and FG wrote the manuscript. All authors contributed to manuscript revision and approval of the submitted version.

Funding

Work in the Grohmann lab was supported by the Deutsche Forschungsgemeinschaft (DFG funding scheme SFB960 TPA7 to DG).

Acknowledgments

We thank all the members of the Grohmann lab for fruitful discussions. Also, we would like to thank all current and previous members of the "archaeal transcription group" (especially Winfried Hausner, Zubeir El Ahmad, and Martin Fenk) for establishment of the genetic system and sequencing protocols and Annett Bellack for providing antibodies. This manuscript has previously appeared online in a preprint, which can be accessed through the DOI: https://doi.org/10.1101/2023.06.15.545098.

Conflict of interest

DG is co-founder of Microbify GmbH. However, there are no commercial interests by the company, or any financial support granted by Microbify GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1241399/full#supplementary-material

References

Albers, S.-V., and Jarrell, K. F. (2015). The archaellum: how archaea swim. *Front. Microbiol.* 6:23. doi: 10.3389/fmicb.2015.00023

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Beeby, M., Ferreira, J. L., Tripp, P., Albers, S.-V., and Mitchell, D. R. (2020). Propulsive nanomachines: the convergent evolution of archaella, flagella and cilia. *FEMS Microbiol. Rev.* 44, 253–304. doi: 10.1093/femsre/fuaa006

Briegel, A., Oikonomou, C. M., Chang, Y., Kjær, A., Huang, A. N., Kim, K. W., et al. (2017). Morphology of the archaellar motor and associated cytoplasmic cone in *Thermococcus kodakarensis*. *EMBO Rep.* 18, 1660–1670. doi: 10.15252/embr.201744070

Canganella, F., Jones, W. J., Gambacorta, A., and Antranikian, G. (1998). *Thermococcus guaymasensis* sp. nov. and *Thermococcus aggregans* sp. nov., two novel thermophilic archaea isolated from the Guaymas Basin hydrothermal vent site. *Int. J. Syst. Bacteriol.* 48, 1181–1185. doi: 10.1099/00207713-48-4-1181

Chaban, B., Ng, S. Y. M., Kanbe, M., Saltzman, I., Nimmo, G., Aizawa, S.-I., et al. (2007). Systematic deletion analyses of the fla genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*. *Mol. Microbiol.* 66, 596–609. doi: 10.1111/j.1365-2958.2007.05913.x

Chaudhury, P., Van Der Does, C., and Albers, S.-V. (2018). Characterization of the ATPase FlaI of the motor complex of the *Pyrococcus furiosus* archaellum and its interactions between the ATP-binding protein FlaH. *PeerJ* 6:e4984. doi: 10.7717/peerj.4984

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Daum, B., Vonck, J., Bellack, A., Chaudhury, P., Reichelt, R., Albers, S.-V., et al. (2017). Structure and in situ organisation of the *Pyrococcus furiosus* archaellum machinery. *Elife* 6:e27470. doi: 10.7554/eLife.27470

Desmond, E., Brochier-Armanet, C., and Gribaldo, S. (2007). Phylogenomics of the archaeal flagellum: rare horizontal gene transfer in a unique motility structure. *BMC Evol. Biol.* 7:106. doi: 10.1186/1471-2148-7-106

Ding, Y., Berezuk, A., Khursigara, C. M., and Jarrell, K. F. (2017a). Bypassing the need for the transcriptional activator EarA through a spontaneous deletion in the BRE portion of the fla operon promoter in *Methanococcus maripaludis*. *Front. Microbiol.* 8:1329. doi: 10.3389/fmicb.2017.01329

Ding, Y., Berezuk, A., Khursigara, C. M., and Jarrell, K. F. (2017b). Phylogenetic distribution of the euryarchaeal archaellum regulator EarA and complementation of a *Methanococcus maripaludis* ΔearA mutant with heterologous earA homologues. *Microbiology* 163, 804–815. doi: 10.1099/mic.0.000464

Ding, Y., Nash, J., Berezuk, A., Khursigara, C. M., Langelaan, D. N., Smith, S. P., et al. (2016). Identification of the first transcriptional activator of an archaellum operon in a euryarchaeon. *Mol. Microbiol.* 102, 54–70. doi: 10.1111/mmi.13444

Dombrowski, N., Williams, T. A., Sun, J., Woodcroft, B. J., Lee, J.-H., Minh, B. Q., et al. (2020). Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11:3939. doi: 10.1038/s41467-020-17408-w

Fiala, G., and Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* 145, 56–61. doi: 10.1007/BF00413027

Gambelli, L., Isupov, M. N., Conners, R., McLaren, M., Bellack, A., Gold, V., et al. (2022). An archaellum filament composed of two alternating subunits. *Nat. Commun.* 13:710. doi: 10.1038/s41467-022-28337-1

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318

Grote, R., Li, L., Tamaoka, J., Kato, C., Horikoshi, K., and Antranikian, G. (1999). *Thermococcus siculi* sp. nov., a novel hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent at the Mid-Okinawa trough. *Extremophiles* 3, 55–62. doi: 10.1007/s007920050099

Grünberger, F., Reichelt, R., Bunk, B., Spröer, C., Overmann, J., Rachel, R., et al. (2019). Next generation DNA-Seq and differential RNA-Seq allow re-annotation of the *Pyrococcus furiosus* DSM 3638 genome and provide insights into archaeal antisense transcription. *Front. Microbiol.* 10:1603. doi: 10.3389/fmicb.2019.01603

Grünberger, F., Reichelt, R., Waege, I., Ned, V., Bronner, K., Kaljanac, M., et al. (2021). CopR, a global regulator of transcription to maintain copper homeostasis in *Pyrococcus furiosus*. *Front. Microbiol.* 11:613532. doi: 10.3389/fmicb.2020.613532

Grünberger, F., Schmid, G., El Ahmad, Z., Fenk, M., Vogl, K., Reichelt, R., et al. (2023). Uncovering the temporal dynamics and regulatory networks of thermal stress response in a hyperthermophile using transcriptomics and proteomics. *mBio*. e02174–23. doi: 10.1128/mbio.02174-23

Haurat, M. F., Figueiredo, A. S., Hoffmann, L., Li, L., Herr, K., Wilson, A. J., et al. (2017). ArnS, a kinase involved in starvation-induced archaellum expression: starvation-induced archaellum expression. *Mol. Microbiol.* 103, 181–194. doi: 10.1111/mmi.13550

Hensley, S. A., Jung, J.-H., Park, C.-S., and Holden, J. F. (2014). *Thermococcus paralvinellae* sp. nov. and *Thermococcus cleftensis* sp. nov. of hyperthermophilic heterotrophs from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* 64, 3655–3659. doi: 10.1099/ijs.0.066100-0

Herzog, B., and Wirth, R. (2012). Swimming behavior of selected species of Archaea. *Appl. Environ. Microbiol.* 78, 1670–1674. doi: 10.1128/AEM.06723-11

Hoffmann, L., Schummer, A., Reimann, J., Haurat, M. F., Wilson, A. J., Beeby, M., et al. (2017). Expanding the archaellum regulatory network - the eukaryotic protein kinases ArnC and ArnD influence motility of *Sulfolobus acidocaldarius*. *Microbiol Open* 6:e00414. doi: 10.1002/mbo3.414

Jarrell, K. F., Albers, S.-V., and Machado, J. N. D. S. (2021). A comprehensive history of motility and archaellation in archaea. *FEMS Microbes* 2:xtab002. doi: 10.1093/femsmc/xtab002

Jarrell, K. F., Ding, Y., Meyer, B. H., Albers, S.-V., Kaminski, L., and Eichler, J. (2014). N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol. Mol. Biol. Rev.* 78, 304–341. doi: 10.1128/MMBR.00052-13

Khan, S., and Scholey, J. M. (2018). Assembly, functions and evolution of archaella, flagella and cilia. *Curr. Biol.* 28, R278–R292. doi: 10.1016/j.cub.2018.01.085

Kreuzer, M., Schmutzler, K., Waege, I., Thomm, M., and Hausner, W. (2013). Genetic engineering of *Pyrococcus furiosus* to use chitin as a carbon source. *BMC Biotechnol.* 13:19. doi: 10.1186/1472-6750-13-9

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S.-V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124. doi: 10.1111/j.1365-2958.2011.07916.x

Lassak, K., Peeters, E., Wróbel, S., and Albers, S.-V. (2013). The one-component system ArnR: a membrane-bound activator of the crenarchaeal archaellum: ArnA activates archaella expression. *Mol. Microbiol.* 88, 125–139. doi: 10.1111/mmi.12173

Lee, S., Cook, D., and Lawrence, M. (2019). plyranges: a grammar of genomic data transformation. *Genome Biol.* 20:4. doi: 10.1186/s13059-018-1597-8

Lee, S. H., Kim, M.-S., Jung, H. C., Lee, J., Lee, J.-H., Lee, H. S., et al. (2015). Screening of a novel strong promoter by RNA sequencing and its application to H2 production in a hyperthermophilic archaeon. *Appl. Microbiol. Biotechnol.* 99, 4085–4092. doi: 10.1007/s00253-015-6444-1

Lewis, D. L., Notey, J. S., Chandrayan, S. K., Loder, A. J., Lipscomb, G. L., Adams, M. W. W., et al. (2015). A mutant ('lab strain') of the hyperthermophilic archaeon *Pyrococcus furiosus*, lacking flagella, has unusual growth physiology. *Extremophiles* 19, 269–281. doi: 10.1007/s00792-014-0712-3

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37, 4572–4574. doi: 10.1093/bioinformatics/btab705

Li, L., Banerjee, A., Bischof, L. F., Maklad, H. R., Hoffmann, L., Henche, A.-L., et al. (2017). Wing phosphorylation is a major functional determinant of the Lrs14-type biofilm and motility regulator AbfR1 in *Sulfolobus acidocaldarius*. *Mol. Microbiol.* 105, 777–793. doi: 10.1111/mmi.13735

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, X., Ma, L., Mei, X., Liu, Y., and Huang, H. (2022). ggmotif: an R package for the extraction and visualization of motifs from MEME software. *PLoS One* 17:e0276979. doi: 10.1371/journal.pone.0276979

Li, J., Yue, L., Li, Z., Zhang, W., Zhang, B., Zhao, F., et al. (2021). aCPSF1 cooperates with terminator U-tract to dictate archaeal transcription termination efficacy. *Elife* 10:e70464. doi: 10.7554/eLife.70464

Liang, R., Robb, F. T., and Onstott, T. C. (2021). Aspartic acid racemization and repair in the survival and recovery of hyperthermophiles after prolonged starvation at high temperature. *FEMS Microbiol. Ecol.* 97:fiab112. doi: 10.1093/femsec/fiab112

Liao, Y., Smyth, G. K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47:e47. doi: 10.1093/nar/gkz114

Maier, L. K., and Marchfelder, A. (2019). It's all about the T: transcription termination in Archaea. *Biochem. Soc. Trans.* 47, 461–468. doi: 10.1042/BST20180557

Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life* 5, 818–840. doi: 10.3390/life5010818

Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., et al. (2014). DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res.* 42, D654–D659. doi: 10.1093/nar/gkt1048

Micorescu, M., Grünberg, S., Franke, A., Cramer, P., Thomm, M., and Bartlett, M. (2008). Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J. Bacteriol.* 190, 157–167. doi: 10.1128/jb.01498-07

Mora, M., Bellack, A., Ugele, M., Hopf, J., and Wirth, R. (2014). The temperature gradient-forming device, an accessory unit for normal light microscopes to study the biology of hyperthermophilic microorganisms. *Appl. Environ. Microbiol.* 80, 4764–4770. doi: 10.1128/AEM.00984-14

Nagahisa, K., Ezaki, S., Fujiwara, S., Imanaka, T., and Takagi, M. (1999). Sequence and transcriptional studies of five clustered flagellin genes from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *FEMS Microbiol. Lett.* 178, 183–190. doi: 10.1111/j.1574-6968.1999.tb13776.x

Näther, D. J., Rachel, R., Wanner, G., and Wirth, R. (2006). Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts. *J. Bacteriol.* 188, 6915–6923. doi: 10.1128/JB.00527-06

Näther-Schindler, D. J., Schopf, S., Bellack, A., Rachel, R., and Wirth, R. (2014). *Pyrococcus furiosus* flagella: biochemical and transcriptional analyses identify the newly detected flaB0 gene to encode the major flagellin. *Front. Microbiol.* 5:695. doi: 10.3389/fmicb.2014.00695

Nuno de Sousa Machado, J., Albers, S.-V., and Daum, B. (2022). Towards elucidating the rotary mechanism of the archaellum machinery. *Front. Microbiol.* 13:848597. doi: 10.3389/fmicb.2022.848597

Ochs, S. M., Thumann, S., Richau, R., Weirauch, M. T., Lowe, T. M., Thomm, M., et al. (2012). Activation of archaeal transcription mediated by recruitment of transcription factor B. *J. Biol. Chem.* 287, 18863–18871. doi: 10.1074/jbc.M112.365742

Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71. doi: 10.1099/ijsem.0.005056

Ouhammouch, M., Langham, G. E., Hausner, W., Simpson, A. J., El-Sayed, N. M. A., and Geiduschek, E. P. (2005). Promoter architecture and response to a positive regulator of archaeal transcription: architecture of an archaeal UAS. *Mol. Microbiol.* 56, 625–637. doi: 10.1111/j.1365-2958.2005.04563.x

Poweleit, N., Ge, P., Nguyen, H. H., Loo, R. R. O., Gunsalus, R. P., and Zhou, Z. H. (2016). CryoEM structure of the *Methanospirillum hungatei* archaellum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nat. Microbiol.* 2:16222. doi: 10.1038/nmicrobiol.2016.222

Pyatibratov, M. G., Syutkin, A. S., Quax, T. E. F., Melnik, T. N., Papke, R. T., Gogarten, J. P., et al. (2020). Interaction of two strongly divergent archaellins stabilizes the structure of the *Halorubrum archaellum*. *Microbiol Open* 9:e1047. doi: 10.1002/mbo3.1047

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Reichelt, R., Ruperti, K. M. A. A., Kreuzer, M., Dexl, S., Thomm, M., and Hausner, W. (2018). The transcriptional regulator TFB-RF1 activates transcription of a putative ABC transporter in *Pyrococcus furiosus*. *Front. Microbiol.* 9:838. doi: 10.3389/fmicb.2018.00838

Reimann, J., Lassak, K., Khadouma, S., Ettema, T. J. G., Yang, N., Driessen, A. J. M., et al. (2012). Regulation of archaella expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in *Sulfolobus acidocaldarius*: ArnA and ArnB regulate archaella expression in *S. acidocaldarius*. *Mol. Microbiol.* 86, 24–36. doi: 10.1111/j.1365-2958.2012.08186.x

Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E., et al. (2020). FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* 5, 545–553. doi: 10.1038/s41564-020-0667-3

Schlesner, M., Miller, A., Streif, S., Staudinger, W. F., Müller, J., Scheffer, B., et al. (2009). Identification of Archaea-specific chemotaxis proteins which interact with the flagellar apparatus. *BMC Microbiol.* 9:56. doi: 10.1186/1471-2180-9-56

Schopf, S., Wanner, G., Rachel, R., and Wirth, R. (2008). An archaeal bi-species biofilm formed by Pyrococcus furiosus and *Methanopyrus kandleri*. *Arch. Microbiol.* 190, 371–377. doi: 10.1007/s00203-008-0371-9

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962

Sievers, F., and Higgins, D. G. (2014). Clustal Omega. *Curr. Protoc. Bioinforma.* 48, 3.13.1–3.13.16. doi: 10.1002/0471250953.bi0313s48

Thomas, N. A., and Jarrell, K. F. (2001). Characterization of flagellum gene families of methanogenic archaea and localization of novel flagellum accessory proteins. *J. Bacteriol.* 183, 7154–7164. doi: 10.1128/JB.183.24.7154-7164.2001

Villain, P., da Cunha, V., Villain, E., Forterre, P., Oberto, J., Catchpole, R., et al. (2021). The hyperthermophilic archaeon *Thermococcus kodakarensis* is resistant to pervasive negative supercoiling activity of DNA gyrase. *Nucleic Acids Res.* 49, 12332–12347. doi: 10.1093/nar/gkab869

Waege, I., Schmid, G., Thumann, S., Thomm, M., and Hausner, W. (2010). Shuttle vector-based transformation system for *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 76, 3308–3313. doi: 10.1128/AEM.01951-09

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647. doi: 10.1093/bioinformatics/btx469

Weiner, A., Schopf, S., Wanner, G., Probst, A., and Wirth, R. (2012). Positive, neutral and negative interactions in cocultures between Pyrococcus furiosus and different methanogenic Archaea. *Microbiol. Insights* 5, 1–10. doi: 10.4137/MBI.S8516

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Wirth, R. (2017). Colonization of black smokers by hyperthermophilic microorganisms. *Trends Microbiol.* 25, 92–99. doi: 10.1016/j.tim.2016.11.002

Wirth, R., Luckner, M., and Wanner, G. (2018). Validation of a hypothesis: colonization of black smokers by hyperthermophilic microorganisms. *Front. Microbiol.* 9:524. doi: 10.3389/fmicb.2018.00524

Yue, L., Li, J., Zhang, B., Qi, L., Li, Z., Zhao, F., et al. (2020). The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3′-end cleavage mode. *Nucleic Acids Res.* 48, 9589–9605. doi: 10.1093/nar/gkaa702

Zeng, X., Zhang, X., and Shao, Z. (2020). Metabolic adaptation to sulfur of hyperthermophilic palaeococcus pacificus DY20341T from Deep-Sea hydrothermal sediments. *Int. J. Mol. Sci.* 21:368. doi: 10.3390/ijms21010368

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q., et al. (2022). ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Brief. Bioinform.* 23:bbac222. doi: 10.1093/bib/bbac222