NAR Genomics and Bioinformatics, 2024, **6**, lqae026
https://doi.org/10.1093/nargab/lqae026
Advance access publication date: 18 March 2024
Standard Article

Evolutionary and functional insights into the Ski2-like helicase family in Archaea: a comparison of Thermococcales ASH-Ski2 and Hel308 activities

Manon Batista[1,†], Petra Langendijk-Genevaux[2,†], Marta Kwapisz[1], Isabelle Canal[1], Duy Khanh Phung[1], Laura Plassart[1], Régine Capeyrou[1], Yann Moalic[3], Mohamed Jebbar[3], Didier Flament[3], Gwennaele Fichant[2], Marie Bouvier[1] and Béatrice Clouet-d'Orval[1,*]

[1]MCD, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse, UT3, Toulouse, France
[2]LMGM, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse, UT3, Toulouse, France
[3]Univ Brest, CNRS, Ifremer, UMR6197 Biologie et Ecologie des Ecosystèmes marins Profonds, F-29280 Plouzané, France

[*]To whom correspondence should be addressed. Tel: +33 5 61 55 69 06; Fax: +33 5 61 33 58 86; Email: Beatrice.Clouet-dOrval@univ-tlse3.fr
[†]The first two authors should be regarded as Joint First Authors.
Present addresses:
Duy Khanh Phung, University College London, Gower Street, Darwin Building, London WC1E 6BT United Kingdom.
Yann Moalic, LabISEN, Yncréa Ouest, 20 rue Cuirassé de Bretagne, F-29200 Brest, France.

Abstract

RNA helicases perform essential housekeeping and regulatory functions in all domains of life by binding and unwinding RNA molecules. The Ski2-like proteins are primordial helicases that play an active role in eukaryotic RNA homeostasis pathways, with multiple homologs having specialized functions. The significance of the expansion and diversity of Ski2-like proteins in Archaea, the third domain of life, has not yet been established. Here, by studying the phylogenetic diversity of Ski2-like helicases among archaeal genomes and the enzymatic activities of those in Thermococcales, we provide further evidence of the function of this protein family in archaeal metabolism of nucleic acids. We show that, in the course of evolution, ASH-Ski2 and Hel308-Ski2, the two main groups of Ski2-like proteins, have diverged in their biological functions. Whereas Hel308 has been shown to mainly act on DNA, we show that ASH-Ski2, previously described to be associated with the 5′-3′ aRNase J exonuclease, acts on RNA by supporting an efficient annealing activity, but also an RNA unwinding with a 3′-5′ polarity. To gain insights into the function of Ski2, we also analyse the transcriptome of *Thermococcus barophilus* ΔASH-Ski2 mutant strain and provide evidence of the importance of ASH-Ski2 in cellular metabolism pathways related to translation.

Introduction

Archaea are critical research models for gaining insights into the origin of the eukaryotic cell and the evolution of life, as well as for studying biogeochemical processes (1,2,3). A recent evolutionary scenario proposes that the eukaryal domain originated within the archaeal domain, arising from the Asgard archaea (4). Archaea represent a diverse phylogenetic group that includes free-living, extremophile, mesophile, symbiont, and opportunistic organisms (5). These micro-organisms share a high significant similarity with informational machineries of Eukarya, such as RNA polymerase, but also regulatory mechanisms with Bacteria, such as operonic organization. It is now important to fill the gap in our knowledge of fundamental post-transcriptional pathways that drive gene expression in Archaea. One of the key regulators at the nexus of multiple pathways of RNA and DNA metabolism are helicases. In here, we decode the relevance of SF2 helicase proteins of the Ski2-like family in the archaeal domain.

The Ski2-like family is a restricted family of superfamily SF2 helicase that includes RNA and DNA helicases. Ski2-like enzymes are reported in eukaryal and archaeal genomes (6,7). Ski2-like helicases display a common helicase core consisting of two RecA-like conserved domains of the SF2 family, a Winged helix domain (WH) and a Ratchet domain. In addition, accessory domains attached to or inserted into the conserved domains modulate their activities and functions. The two RecA-like domains which form a ring around the 3′ end of RNA or DNA substrates, contain up to 12 conserved sequence motifs involved in nucleic acid transaction, in nucleotide tri-phosphate binding and hydrolysis, and in coordination of unwinding and annealing activities (8). The reported Ski2-like helicases possess an unwinding activity with 3′ to 5′ directionality (7). The translocation mechanism of Ski2-like helicase along RNA or DNA has been proposed based on resolved structures of the archaeal Hel308 helicase from *Archaeoglobus fulgidus* (9). However, current knowledge on Ski2-like enzymes mostly comes from model organisms belonging to the eukaryal domain. Four Ski2-like RNA helicases (Mtr4, Ski2, Brr2 and Slh1/Rqt2), all part of large multiprotein complexes, play critical roles in nuclear and cytoplasmic RNA decay, ribosome-associated quality control and pre-mRNA splicing. These helicases hydrolyse ATP to power processive RNA unwinding with 3′-5′ polarity (7). Mtr4 and Ski2 RNA helicases that share biochemical properties are

Received: May 10, 2023. **Revised:** February 19, 2024. **Editorial Decision:** February 26, 2024. **Accepted:** February 27, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

associated with the nuclear or cytoplasmic RNA exosome, respectively. They are hubs that shape the transcriptome by orchestrating exosome-dependent 3′-5′ RNA decay. To facilitate the recruitment of RNA substrates by the RNA exosome, they uncoil secondary RNA structures and remodel RNP complexes to free the RNA 3′ end (10,11,12,13). Mtr4 interacts directly with the Rrp6 subunits of the nuclear RNA exosome and various adaptor factors to form mutually exclusive complexes named TRAMP, PAXT and NEXT to target specific RNA substrates (11,14,15,16). Ski2 is part of the Ski complex, composed of the Ski3, Ski8 and Ski7 proteins, that channels the ribosome-bound RNA to the RNA exosome to trigger mRNA degradation in the cytoplasm (12,17,18,19,20). Remarkably, Brr2 and Slh1/Rqt2 are particularly large RNA helicases with duplicated Ski2-like domains (7). Brr2 is an integral component of the spliceosomal U5 snRNP and of the preassembled U4/U6.U5 tri snRNP, essential for activating the core splicing machinery (21,22,23). Finally, the Ski2-like helicase 1 (Slh1, also named Rqt2) was shown to be implicated in the Ribosome Quality Control pathway (RQC) by being part of the RQC-trigger complex (RQT) that degrades the stalled nascent chain during translation arrests (24,25). Recent studies propose that Slh1 applies a pulling force on the mRNA resulting in ribosomal subunit dissociation (26). In eukaryotes, some Ski2-like enzymes are DNA helicases such as yeast Mer3/mammalian HFM1, involved in homologous recombination (27,28) and eukaryotic PolQ-like helicase involved in double-strand break repair (29).

In Archaea, Hel308 is the best-known member of the Ski2-like family, originally named Hjm for 'helix junction migration' and identified as a structure-specific DNA helicase. Studies have reported that Hel308 promotes genome stability in the context of DNA replication (30,31,32). Homologues designated as PolQ/HelQ are present in eukaryotes and have been found to play a role in DNA metabolism (33,34). Reports indicate that Hel308/PolQ has a preference for targeting replication forks in *Pyrococcus furiosus*, *Methanothermobacter thermautotrophicus*, *Sulfolobus tokodaii*, *Sulfolobus islandicus* and humans (31,32,34,35). More specifically, Hel308/Hjm from *S. tokodaii* and *S. islandicus* interact with the Hjc endonuclease and the PINA ATPase, respectively. Both of these enzymes are involved in resolving Holliday junctions (35,36). Recent studies have suggested that Hel308/Hjm from *M. thermautotrophicus* not only possesses ATP-dependant DNA helicase activity, but also autonomously anneals single-stranded DNA without ATP (37). More recently, a member of another group of the Ski2-like family was reported to be part of RNA metabolic pathways (38). This factor was initially named ASH-Ski2 for Archaeal Specific Helicase Ski2 (6,38), but is also known as Eta for Euryarchaeal termination activity (39). In Thermococcales, ASH-Ski2 was proposed to be part of an RNA degrading complex by interacting with the 5′-3′ exoribonuclease aRNase J but also to be involved in the rescue of transcriptional arrest (38,40). Nevertheless, many questions remain open on the roles of Ski2-like helicases in archaeal DNA and RNA metabolism.

To address the significance of archaeal Ski2-like helicases, we studied their diversity and distribution among archaeal genomes, and we deciphered the biochemical specificities of ASH-Ski2 and Hel308 from Thermococcales. This study shows that the enzyme activities of ASH-Ski2 and Hel308, the major groups of archaeal Ski2-like proteins, are divergent. Hel308 is ubiquitously conserved among the archaeal domain whereas ASH-Ski2 is restricted to the euryarchaeal clade. By comparing the *in vitro* enzymatic activities of both ASH-Ski2 and Hel308 from the *Pyrococcus abyssi* archaeon, we confirmed that indeed Hel308 mainly acts on DNA substrates as previously proposed in *P. furiosus* and in Sulfolobales (31,32,35,36). In addition, we provide the first evidences that ASH-Ski2 is an RNA-helicase that has an unwinding activity with a 3′-5′ polarity and an efficient annealing activity of RNA duplexes. Finally, we analysed the transcriptome of a ΔASH-Ski2 strain to obtain the first indications of the impact of ASH-Ski2 on cellular metabolism, notably at the level of the translation process. These results are a milestone towards the role and importance of the Ski2 helicase family in the RNA metabolism of archaeal organisms.

Materials and methods

Building a Ski2-type helicase dataset

219 non-redundant completely sequenced and well annotated archaeal genomes were retrieved *via* the API from the ENA repository at EMBL-EBI. To increase Archaea's diversity, we also retrieved one Asgard proteome from the Proteome section of UniProt (*Candidatus* Prometheoarchaeum syntrophicum), and 74 Asgard genomes for which the NCBI RefSeq's genome assembly assessment is annotated as 'full' (34 Lokiarchaeota, 27 Thorarchaeota, 12 Heimdallarchaeota and one Odinarchaeota). However, their sequencing and annotation is of lower quality than the other archaeal genomes. Ten representative eukaryotic genomes were also retrieved from the Proteome section of UniProt (*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Danio rerio*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *Gallus gallus*, *Giardia intestinalis*, *Homo sapiens*, *Plasmodium falciparum*, *Saccharomyces cerevisiae*) for which the Complete Proteome Detector (CPD), which estimates the completeness and quality of the proteome, is labelled to 'standard' and the NCBI RefSeq's assessment of genome assembly representation is annotated as 'full'. These complete genomes or proteomes were processed by a set of *perl* programs into in house mySQL databases. We used the *RPS-BLAST* program to annotate protein sequences according to the conserved domain database downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (41). The two previously identified Ski2-like archaeal subfamilies ASH-Ski2 and Hel308-Ski2 (6) are characterized by the COG1202 (38) and the COG1204, respectively. The two well-studied human Mtr4 helicase (SwissProt: P42285) and yeast Ski2 helicase (SwissProt: P35207) are both characterized by the COG4581. RPS-BLAST hits with one of these three COG showing a score >60 and an alignment covering of at least 60% of the COG were retrieved. An initial set of 519 archaeal proteins and 70 eukaryal proteins was first obtained. After careful examination of these sequences, we removed 53 archaeal sequences that were partial or had partially conserved motifs, or that did not have all the characteristic motifs of the sub-families. We ended up with a set of 466 sequences of archaea, 179 obtained with the COG1202 (ASH-Ski2), 287 with the COG1204 (Hel308-Ski2) and none with the COG4581 (Supplementary Table S1). The 70 eukaryal sequences include 44 proteins annotated with the COG1204 and 26 with the COG4581. As expected, no hit with the COG1202 was obtained since the ASH-Ski2 subfamily was shown to be specific to Archaea (6).

Helicase phylogenetic tree construction

Brr2 eukaryotic helicases are large proteins that contain two Ski2-like domains (Figure 1). However, only the first helicase domain has been shown to have functional helicase activity (42). Therefore, to compare the eukaryotic Brr2 helicases with the other Ski2-like helicases, we split their sequences and retained for the alignment only the N-terminal region of the protein encompassing the first Ski2-like domain, *i.e.* from the beginning of the protein to the end of the first Sec63 domain. The 536 Ski2-like helicases were aligned with *Muscle v5*.1 (https://drive5.com/muscle5/; (43)). In order to eliminate the variability of the N- and C-term regions of the proteins, the alignment has been edited with Jalview (http://www.jalview.org; (44)) to extract the common RecA1-RecA2 core of conserved residues from the Q motif to the motif VI. In order to improve the quality of the sequence alignment by removing spurious homologies while retaining informative data, we used the *divvier* method (45) with the divvying filtering option (parameters: *-divvygap -mincol 4*). This trimmed alignment was used to infer the phylogenetic tree. The amino acid substitution model LG4X + I was selected by *modeltest-ng* (46) as being the one that best fit the data. The tree was computed with *iq-tree* software (47). The ultra-fast bootstrap approximation (*-bb 1000*) and SH approximate likelihood ratio test (*-alrt 1000*) were used to estimate the branch supports. The tree was annotated and visualized with the online tool Interactive Tree Of Life v6 (iTOL, https://itol.embl.de; (48)).

Archaeal species tree construction

To construct the archaeal species tree, we used the 122 markers that have been identified as reliable for phylogenetic inference (49). The set of 122 protein markers is characterized by HMM profiles from the Pfam (v27) and the TIGRFAMs (v15.0) databases. For each genome included in this study, the proteins were identified by using each profile entry as query with the *hmmsearch* program from the HMMER3 package with the *–cut_tc* (trusted cutoff) parameter (http://hmmer.org/; (50)). The *hmmsearch* output domain file was parsed to extract for each genome and for each HMM profile the best protein hit. However, despite being annotated as 'full' in UniProt, some Asgard genomes are still partial. To retain a species, we required that at least 75% of the markers (90 markers) could be retrieved. Only 42 out of the 75 Asgard genomes met our criterion, so the other 33 were excluded from the species tree reconstruction which was performed by considering 259 archaeal species. Protein alignments with HMM profiles were merged for each marker. We thus obtained 122 sequence alignments. The columns of the alignments that had a high deletion frequency were removed with *trimal* (*-gt 0.1*) (51). The quality of the alignments was estimated using the *t-coffee* transitive consistency score (TCS) (52). The analysis of the results obtained on each alignment allowed (i) to eliminate sequences with outlier TCS values and (ii) to discard two alignments (PF04104.9 and PF01990.12) with a low overall TCS value (TCS < 65). The resulting 120 marker alignments were concatenated and if for a given species, the marker protein sequence is missing, it is reintroduced into the alignment as gaps. The model LG + I + G4 + F was selected with *modeltest-ng* and the tree was inferred with *iq-tree* and branch support values were determined using ultra-fast bootstrap approximation (*-bb 1000*) and the SH approximate likelihood ratio test (*-alrt 1000*). The tree was rooted on DPANN Archaea according to (53). We compared the topology of the archaeal tree with those previously published by Parks *et al.* (49) and Williams *et al.* (53), respectively. We identified the same phylogenetic relationships as previously observed between the major taxonomic groups of archaea, despite using a different number of species genomes for the tree reconstruction.

ASH and Hel308 phylogenetic tree construction

ASH and Hel308 sequences have been aligned independently to further analyse their evolution. The alignments were performed with *Muscle v5.1* and the aligned positions were filtered with the *divvier* method as previously described for the full helicase sample. It is often difficult to infer a well-supported phylogenetic tree for individual protein families, even with the best-fit substitution model, as multiple sequence alignments usually contain insufficient phylogenetic signal due to their often too short length. To circumvent these difficulties, we used GeneRax, a maximum likelihood species-tree-aware phylogenetic inference software (54), which can infer a rooted family tree directly from the multiple sequence alignment and a rooted species tree or optimize the topology of a previously computed family tree according to the species tree. We tested both approaches and more relevant gene tree topologies, i.e. with a higher number of proteins (genes) predicted to have been inherited vertically, were obtained by providing GeneRax with a starting gene tree. Hence, for ASH-Ski2 and Hel308-Ski2, we first computed a starting gene tree by selecting the amino acid substitution model that best fit the data with *modeltest-ng* (46) and inferred phylogenetic trees using *iq-tree* software (47). The best model LG + I + G4 + F was selected for ASH-Ski2 and LG + R4 + F for the Hel308-Ski2 dataset. The ultra-fast bootstrap approximation (*-bb 1000*) and the SH approximate likelihood ratio test (*-alrt 1000*) were used to estimate the branch supports. Subsequently, the archaeal species tree, the multiple alignment and protein tree of ASH-Ski2 or Hel308-Ski2 were provided to GeneRax. The tree optimization was performed with Subtree Prune and Regraft (SPR) moves (*–strategy SPR*) with the recommended value of maximum radius (*-max-spr-radius 5*) used for SPR moves in the tree search. The reconciliation of the family tree with the species tree is performed by assuming that genes evolve through gene duplication, gene loss, horizontal gene transfer, and speciation events, *i.e.* the genes are vertically inherited (*–rec-model UndatedDTL*). The substitution models used are those determined above (*–subst_model LG + I + G4 + F* for ASH-Ski2 and *–subst_model LG + R4 + F* for Hel308-Ski2).

Construction of pET11b expression vectors

The Supplementary Table S2 summarizes the oligonucleotides used to construct pET11b expression vectors. All constructions were obtained by assembling PCR fragments using InFusion® cloning kit (Takara). Using appropriate sets of oligonucleotides, the pET11b (untagged protein) vectors were linearized by PCR amplification with the PrimeSTAR Max DNA polymerase (Takara), and the coding sequence of *P. abyssi* ASH-Ski2 (PAB2313) and Hel308 (PAB0592) were amplified from genomic DNA with the Phusion High-Fidelity DNA polymerase (ThermoFisherScientific). The pET11b vector expressing the ASH-Ski2$_{K245A}$ variant was generated by site-directed mutagenesis of its wildtype counterpart with appropriate sets of oligonucleotides using the QuikChange II

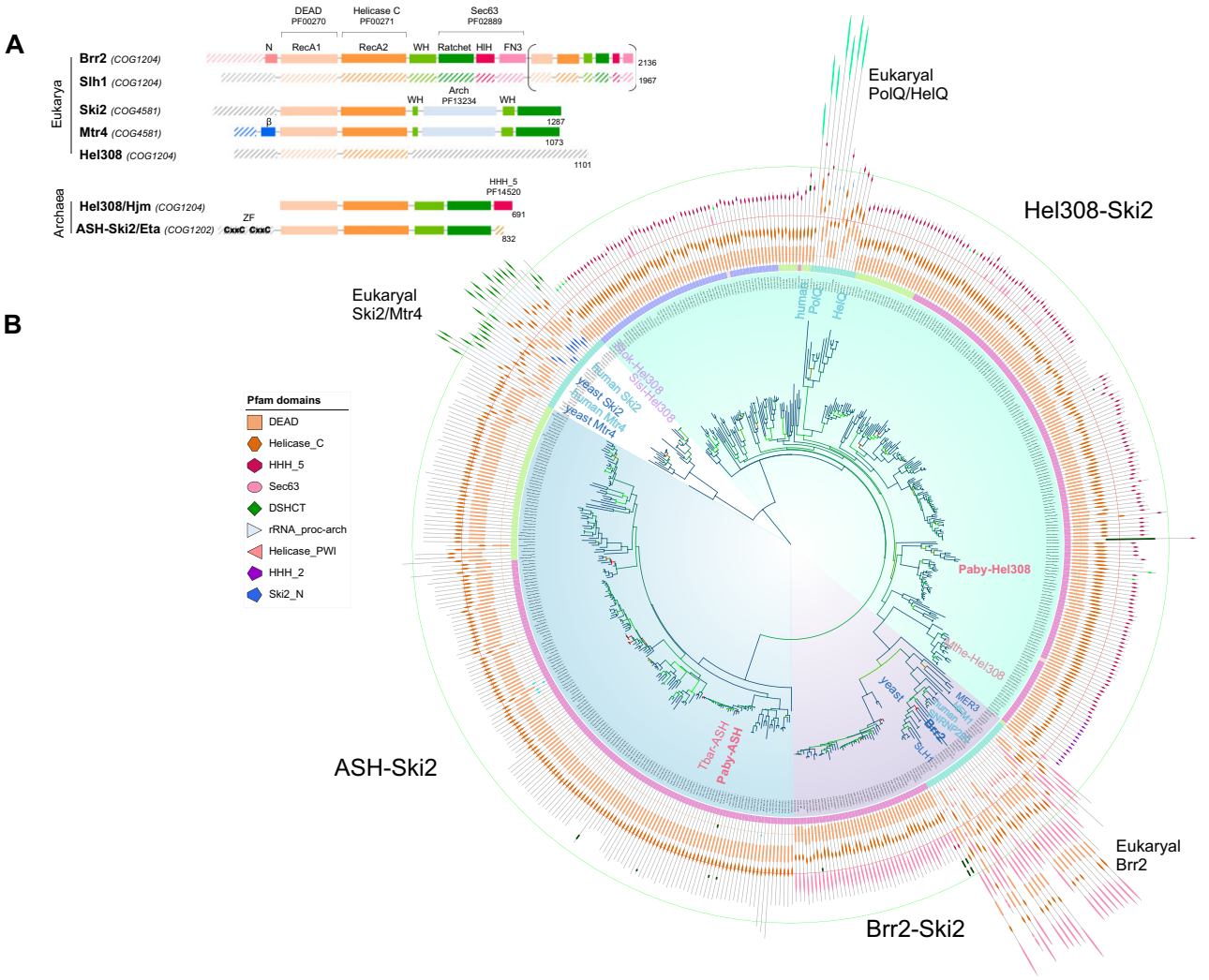


Figure 1. (**A**) Domain architecture of eukaryal and archaeal Ski2-like helicase family members. The human Brr2 (SNRNP200), *Saccharomyces cerevisiae* Slh1 (G9365), Ski2 (L8084.17) and Mtr4 (J1158) and *Pyrococcus furiosus* Hel308-Ski2/Hjm (PF0677) architectures are compared to the one of *Pyrococcus abyssi* ASH-Ski2 (PABY2313). The conserved SF2 core composed of the RecA1 and RecA2 domains are in orange shades. The Winged Helix (WH) and Ratchet domains, outlining the Ski2-like family, are in green shades. The Helix hairpin Helix (HhH) and Helix loop Helix (HIH) found in Brr2 and Hel308-Ski2 and the Fibronectin type III domain (FN3) only present in Brr2 are in pink shades. For Brr2, the Ratchet, HhH and FN3 domains are grouped together in the Sec63 domain. In Mtr4 and Ski2, the Arch domain (PF13234) in light blue is inserted into the WH domain. Specific accessory domains are shown in blue, pink and grey shades. The β hairpin (β) and the Zinc-Finger (ZF) domains typical of Mtr4 and ASH-Ski2, respectively are in blue and grey. Three-dimensional structures have been solved for Brr2, Ski2, Mtr4, Hel308-Ski2 and ASH-Ski2 impaired for its N-terminal domain. The COG for each member is indicated. (**B**) Phylogenetic tree and domain organization of archaeal and eukaryal Ski2-like helicases. The tree is rooted to the branch that separates the subtrees containing Ski2 and Mtr4 eukaryotic sequences. The branches are coloured according to their ultra-fast bootstrap value using a colour gradient from red (bootstrap value of 0) to blue (bootstrap value of 100) with flashy green at midpoint values. The taxonomic origin of the sequences is shown in the first outer ring: pink for Euryarchaeota, purple for TACK-group, light green for Asgard, orange for DPANN, light orange for environmental samples and turquoise for eukaryotes. To highlight the different Ski2-like subfamilies, their corresponding subtrees are coloured: light green for Hel308-Ski2 sequences, light blue for ASH-Ski2 proteins and light purple for Brr2-Ski2 homologs. The Pfam motif architecture for each member is specified at the circumference of the tree. Their colour code is given in the legend panel 'Pfam domains'. The location of reference sequences as well as those of *P. abyssi* and *T. barophilus* have been highlighted. The sequence identifiers for each subtree are in Supplementary Table S1.

XL Kit (Stratagene). The pET11b vector expressing the truncated ASH-Ski2$_{\Delta N}$ was constructed by reverse PCR on the pET11b::ASH-Ski2$_{WT}$ using specific phosphorylated oligonucleotides and by DNA ligation (T4 DNA ligase).

Purification of ASH-Ski2 and Hel308 recombinant proteins from *P. abyssi*

E. coli BL21-codon+ (DE3) cells freshly transformed with pET11b::ASH-Ski2, pET11b::ASH-Ski2$_{\Delta N}$, pET11b::ASH-Ski2$_{K245A}$ and pET11b::Hel308 vectors were grown in 400 ml of LB medium at 37°C. Protein production was induced at OD_{600nm} 0.6 with 0.15 mM IPTG. After 3h of induction at 30°C, the cells were collected, suspended in 10 ml of lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol) supplemented with 1 mg.ml^{-1} of lysozyme and a mix of EDTA-free protease inhibitor (cOmplete™, Roche, Merck KGaA, Darmstadt, Germany), and lysed by sonication (4x [5 × 10 sec], 50% cycle, VibraCell Biolock Scientific). The cleared extracts, obtained by centrifuging the crude extract (20 000 g, 4°C, 20 min), were treated with a mix of RNase A (20 μg ml^{-1}), RNase T1 (1 U μl^{-1}) and DNase I

(20 μg ml^{-1}) containing 10 mM of $MgCl_2$ for 30 min at 37°C. After a heating step at 70°C for 20 min, the extracts were further clarified by centrifugation (20 000 g, 4°C, 20 min). The recombinant proteins were purified from the soluble fractions to near homogeneity using FPLC (Fast Protein Liquid Chromatography, Äkta-purifier10, GE-Healthcare) by a heparin column (Heparin FF) and ion exchange column (SP-HP for ASH-Ski2, ASH-Ski2$_{\Delta N}$ and ASH-Ski2$_{K245A}$, and Q-HP for Hel308) with a linear gradient of NaCl (300 mM to 1 M). All the recombinant proteins were then dialyzed in a buffer containing 20 mM HEPES pH 6.8, 300 mM NaCl, 10% Glycerol. Using Bradford protein assays, protein concentrations of ASH-Ski2, ASH-Ski2 $_{K245A}$, ASH-Ski2$_{\Delta N}$ and Hel308 were determined to be 15, 5, 2 and 10 μM, respectively. The oligomerization state of ASH-Ski2 was determined by size exclusion chromatography. 5 μM of the purified protein was loaded on a size-exclusion Superdex 200 Increase 10/300 GL column in 20 mM HEPES pH 7.5, 300 mM NaCl, 5% Glycerol. The fractions were analysed by Coomassie-blue SDS-PAGE and Western blotting.

Preparation of radiolabelled nucleic acid substrate

The 26-nt RNA (RNA_{26}) and the DNA (DNA_{26}, DNA_{31}, DNA_{50} and DNA_{59}) oligonucleotides were synthesized by Eurofins Genomics. The 50-nt RNA substrate (RNA_{50}) was obtained by *in vitro* transcription from a PCR fragment where DNA_{50} was fused to the T7 promoter using the MEGAscript kit (Ambion). The DNA and RNA substrates were 5′-end radiolabelled using T4 polynucleotide kinase and γ-^{32}P-ATP. To prepare nucleic acid duplexes, the radiolabelled short DNA or RNA oligonucleotides were mixed with complementary unlabelled DNA or RNA strands at a 1:1 molar ratio (100 nM each), incubated for 5 min at 95°C in 1× SSC buffer (composition), and then slowly cooled at room temperature. The nucleotide sequences of all the substrates used in this study are given in (Supplementary Table S3).

Nucleic acid binding assay

Double filtration binding assays were performed with range of protein concentrations from 0 to 200 nM and 0.5 nM of γ-^{32}P-labelled RNA or DNA substrates using a Slot blot device (Amersham Biosciences). The protein was preincubated for 10 min at 65°C in 25 mM Tris–HCl pH 7.5, 50 mM NaAc pH 5.2, 5 mM $MgCl_2$ and 2.5 mM β-Mercaptoethanol. Upon adding the substrate, the reactions were incubated for 15 min at 30°C. Free nucleic acids were separated from nucleoprotein complexes on double filtration systems using Nylon and Nitrocellulose membranes (AmershamTM Hydond-N and Protran, respectively). Radioactive signals were measured using a PhosphorImager device and quantified with MultiGauge software. The apparent dissociation constants K_D were calculated using GraphPad Prism 7 software.

ATP hydrolysis assay

250 nM of recombinant protein were mixed with *P. abyssi* total RNA or genomic DNA in a 50 mM HEPES pH 7.5, 50 mM KCl, 1 mM $MgCl_2$, 2 mM DTT buffer and preincubated for 10 min at 65°C. 2 mM ATP and 0.85 μCi γ-^{32}P-ATP were added at the 0-time point. The kinetic process was performed at 65°C. At the indicated time, aliquots were spotted directly onto the TLC plate (PEI-cellulose, Macherey Nagel SAS, Hoerdt, France). TLC were developed with 0.25 M KH_2PO_4. Radioactive signals were measured as previously and the plots were derived using GraphPad Prism 7 software. Total RNA or genomic DNA from *P. abyssi* were prepared according to (55) and (56), respectively.

Helicase unwinding assay

The unwinding assays were done with 250 nM of protein, 5 nM of γ-^{32}P-labeled nucleic acid duplex and a 200-fold excess of the unlabelled oligo trap (1 μM). The protein was preincubated separately for 5 min at 65°C in 25 mM Tris–HCl pH 8, 50 mM NaAc, 2.5 mM β-Mercaptoethanol, 5 mM $MgCl_2$, 5 mM ATP. After addition of the recombinant protein, the reaction mixtures were incubated at 65°C for the indicated times and then quenched with 0.1% SDS, 8 mM EDTA, 0.1 mg ml^{-1} Proteinase K, 0.02% Bromophenol blue, and 4% glycerol. The reaction products were separated on a native 8% polyacrylamide gel (1× TBE, 0.1% SDS) by electrophoresis in 1× TBE (200 V, 90 min). Radioactive signals were measured using a PhosphorImager device and quantified with MultiGauge software. All assays were repeated at least three times. The plots were derived using GraphPad Prism 7 software.

Strand-annealing assay

5 nM of radiolabelled substrates and 250 nM of recombinant protein were preincubated separately for 5 min at 65°C in 25 mM Tris–HCl pH 8, 50 mM NaAc and 2.5 mM β-Mercaptoethanol. The reactions were started by mixing the protein and nucleic acid samples. After incubation at 65°C, samples of 5 μl were withdrawn at the indicated time points. The reactions were quenched and analysed as above. All assays were independently repeated at least three times.

Transcriptomic analysis

The strains of *Thermococcus barophilus* used for transcriptomic were UBOCC-M-3300 and UBOCC-M-3315, respectively harbouring the genotypes Δ*TERMP_RS02570* (old locus tag *TERMP_00517)* (Wild-Type reference) and Δ*TERMP_RS02570* Δ*ASH-Ski2*. The chromosomal copy of *TERMP_RS08745* (old locus tag : *TERMP_01768*, encoding *Tba*-ASH-Ski2) was deleted by the pop-in/pop-out method to generate the *T. barophilus* ΔASH-Ski2 strain as described in (38). Biomass productions were performed in standard growth conditions (TRM with sulphur, anaerobia, 85°C under atmospheric pressure). Total RNAs were extracted from three independent exponentially grown cultures (50 ml of culture containing by means 10^7 of cells, 6h after inoculation from over-night stationary culture) and the mean quantity of total RNAs was 280 ng.μl^{-1} (RIN = 8.8).

Samples were sequenced by Genewiz through two independent batches of 30 samples in total using a HiSeq Illumina sequencer (paired-end sequencing, 150 bp reads, ∼75M reads/sample). The quality of each raw sequencing file (fastq) was verified with FastQC. All files were aligned to the reference *T. barophilus* genome (GCA_000151105.2) using STAR aligner (57). The read count per sample was computed using HT-seq count (58). The raw count table was cleaned for lowly expressed genes, genes with an averaged raw read count per sample higher or equal to five were kept. Based on multivariate data exploration: PCA and hierarchical clustering with Euclidian distance, the sample WT2 was identified as outlier and removed from the analysis. Differential analysis was then applied to compare three mutant samples against

five WT samples using DESeq2 Bioconductor R-package (59). DESeq2 normalization was applied on the raw read counts, taking account of the type and the batch effect, and $\log_2$ fold change ($\log_2$FC) values were computed for each gene. *P*-values for the significance of each fold change were then computed and adjusted for multiple testing. Single gene analysis was performed using the RT-qPCR method. cDNA produced on total RNA template with reverse transcriptase (Superscript VI, Invitrogene) and random hexamers was quantified by real-time qPCR using iTaq universal SYBR Green Supermix (Bio-Rad) and the ViiA7 AB Applied Biosystems (Life Technologies). Primer pairs used for amplification are available upon request. Signals were analysed with Quant Studio Real Time PCR Software v1.1 and were normalized to TERMP_RS03110 (old locus tag TERMP_00623; NTPase). Error bars correspond to standard deviations of three independent cultures for each strain. In order to apply a Gene Ontology Analysis with R-packages we decided to generate our own OrgDb package for *T. barophilus*. The two databases Quickgo and UniProt were merged cleaned and organized according to orgDb requirements. The AnnotationForge Bioconductor R-package (60) was used to create a custom in-house R-package. In this database, the type of ontology was missing (Molecular function (MF), biological process (BP) or cellular component (CC)). To complete our in-house R-package ontology information, other well-annotated genomes (human, mouse, drosophila, *C. elegans*, *S. cerevisiae*) databases were used in order to select the type of ontology for common GO terms. Ontology types for 961 GO terms were found. Gene Ontology was then performed on the significantly differential genes (adjusted *P*-values < 0.05) using ClusterProfiler Bioconductor R-package (61). Transcripts Per Kilobase Million (TPM) were mapped onto the genome of *T. barophilus* (Refseq NC_014804.1) with Anvi'o (v7) (62) as for *Thermococcus piezophilus* RNA-seq data processing (63).

Results

Comprehensive search for Ski2-like protein members in archaeal genomes

In contrast to the roles of eukaryotic Ski2-like helicases in RNA homeostasis, functions and reports of archaeal Ski2-like helicase family members are still scarce (6). Among the SF2 helicase families, the Ski2-like proteins are defined by unique accessory domains that decorate the helicase core and provide additional functionalities (8,64) (Figure 1A). Combined, Ski2-like structures have demonstrated that the molecular 'core' of all Ski2-like helicases is a ring-like four domain assembly of two adjacent RecA-like domains corresponding to the 'DEAD' (RecA1) and 'Helicase C' (RecA2) Pfam domains, a winged helix domain and a ratchet domain (7,65) (Figure 1A). We achieved an in-depth phylogenomic analysis to get insights into the significance of the Ski2-like helicase family in Archaea. Our study included 294 archaeal genomes, of which 219 non-redundant genomes were fully sequenced and well annotated. To ensure a coverage of archaeal diversity, we also included 75 Asgard genomes, which are partially annotated. To get further insight into the evolutionary relationships among members of this family, we added ten eukaryotic genomes of representative organisms to highlight the relationships between eukaryotic and archaeal Ski2-like members.

We performed similarity searches against the COG1202, COG1204 and COG4581 profiles from the Conserved Domain Database that cover archaeal ASH-Ski2 and Hel308-Ski2 helicases as well as eukaryal Mtr4 and Ski2 helicases, respectively. We collected 536 members, including 70 belonging to eukaryotes (Supplementary Table S1). A maximum-likelihood Ski2-like phylogenetic tree was inferred from the multiple sequence alignment of the sole SF2 helicase core region composed of the RecA1-RecA2 domains in order to eliminate the variability of the N-terminal and C-terminal regions of the Ski2-like proteins. Using the Pfam domain organization and the tree topology, we can distinguish four distinct groups (Figure 1B). We named each group based on previously reported features of some members and their location on the tree (6,7).

Two groups corresponding to two subtrees, well supported by bootstrap values of 100%, can be clearly identified. The first one, which was used to root the tree, corresponds to the unique group exclusively composed of eukaryal sequences that segregates into two subtrees corresponding to Mtr4 and Ski2 proteins respectively and that is characterized by the COG4581 (Figure 1B, in white). The second group, at the opposite end of the tree, that is monophyletic and specific of the archaeal domain, corresponds to the ASH-Ski2 proteins and is characterized by the COG1202 (Figure 1B, in blue). The remaining sequences are found under a same ancestral node and all possess the COG1204. According to the domain organization and the tree topology, two distinct sub-groups emerged. One sub-group sharing a common domain organization, is not monophyletic. This subgroup includes the archaeal Hel308 sequences from *M. thermautotrophicus* (MtheA01.AAB85310.1; SP: O26901), *S. tokodaii* (StokA01.HJM; SP: Q974S1) and *S. islandicus* (SislA01.ACP34435.1; SP: C3MKR0) as well as the eukaryotic PolQ and HelQ sequences (Figure 1B, in green). This group was accordingly termed, Hel308-Ski2. Its topology is not as well resolved and it harbours a large number of short deep branches. Remarkably, we noticed that the Hel308-Ski2 eukaryotic sequences branch with the Asgard sequences highlighting the close relationship between Asgard and Eukarya (66). The second sub-group includes eukaryotic Brr2 sequences and forms a well-supported subtree (100% SH-aLRT bootstrap support). This sub-group was termed Brr2-Ski2, due to the identification of the sec63 Pfam domain in both the archaeal and eukaryotic sequences (Figure 1B, in violet). We observed that the Brr2-Ski2 subtree emerged from the Hel308-Ski2 group.

The taxonomic distribution of the retrieved archaeal sequences within each of the three groups indicates that most archaeal genomes encode Hel308 proteins (Figure 2). The absence of Hel308 in 18 of the 42 Asgard genomes, including 13 of the 15 Lokiarchaeota genomes, remains questionable given the lower quality and incompleteness of these genomes. Regarding the ASH-Ski2 group, its members appear to be restricted to most euryarchaeal groups, except for the Archeoglobales and Thermoplasmata as previously observed (38) (Figure 2). In this study, we identified ASH-Ski2 members in 38 out of 42 Asgard genomes providing evidence for the presence of ASH-Ski2 within the Asgard group. Finally, the Brr2-Ski2 members are restricted to Halobacteria, two genomes of Methanocellales and one Asgard genome (*Candidatus Heimdallarchaeota archaeon LC_3*) (Figure 2).

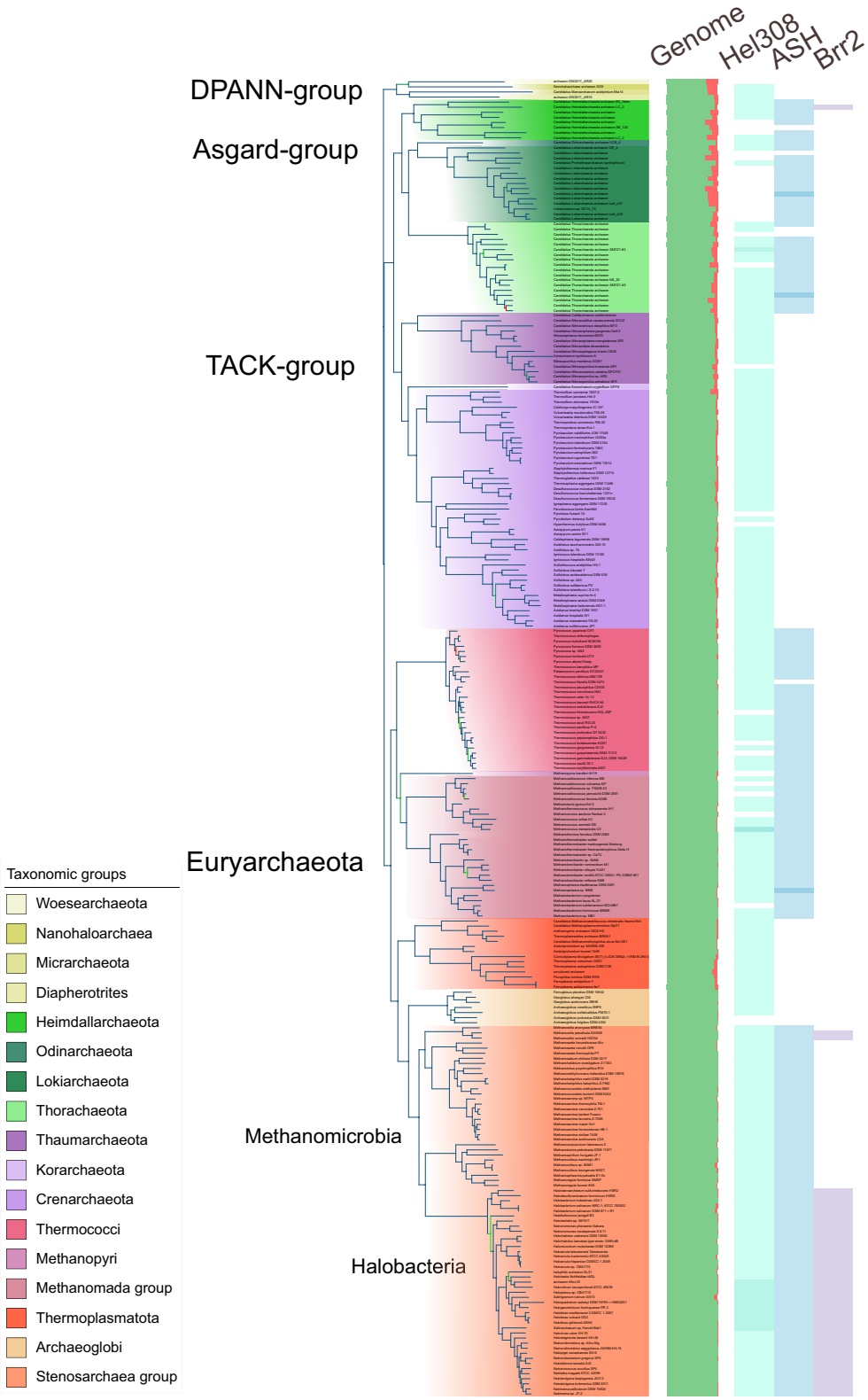


Figure 2. Taxonomic distribution of Hel308-Ski2, ASH-Ski2 and Brr2-Ski2. The species tree of the archaeal genomes was deduced from sets of 120 concatenated sequence alignments. The branches are coloured according to their ultra-fast bootstrap value using the same colour gradient as above. The tree was rooted on DPANN Archaea according to (53). The taxonomic colour code is given in the legend panel 'taxonomic groups' in the tree order, from top to bottom. An indicative genome completeness, based on the presence (green) or absence (red) of the 120 markers used for the tree construction, is shown as a histogram adjacent to the species tree. Occurrences of each subfamily are plotted using the same colour code as above. Darker shades indicate multiple copies.

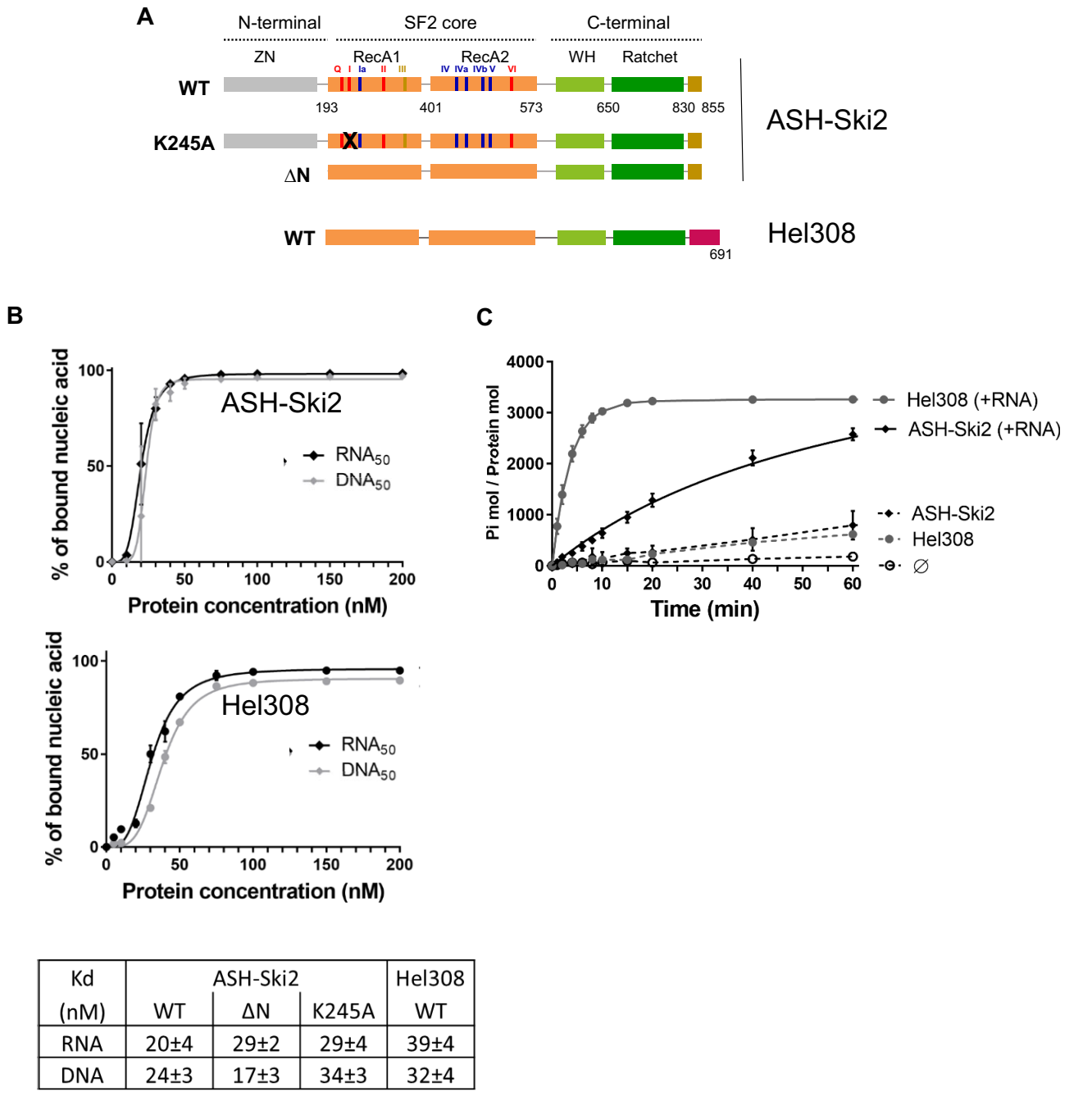
To go further into the evolutionary routes undertaken by both ASH-Ski2 and Hel308-Ski2 sequences, we inferred two maximum-likelihood phylogenetic trees using GeneRax (54) (Supplementary Figure S1). The tree topologies were optimized using the species tree, gene duplication, gene loss, horizontal gene transfer (HGT) and vertical inheritance (called speciation in GeneRax). We compared the ASH-Ski2 and Hel308-Ski2 tree topologies with the archaeal species tree (Supplementary Figure S1). The tips of the reconciled protein tree are linked to the corresponding species of the tree. Incongruences between both trees are revealed by the crossing of the lines connecting the species to the corresponding protein. For the reconciled ASH-Ski2 tree (Supplementary Figure S1A), GeneRax predicted 141 events of vertical inheritance, three duplication events (two being recent duplications corresponding to the presence of duplicated genes in two current species), 25 HGTs and 29 loss events. Only a few incongruences are observed within taxonomic groups but none between groups (Supplementary Figure S1A). For the reconciled Hel308-Ski2 tree (Supplementary Figure S1B), GeneRax has predicted 206 events of vertical inheritance, 4 duplications events (two being recent duplications corresponding to the presence of duplicated genes in two current species, and an older one encompassing mainly Haloferacales genomes (Supplementary Figure S1B), 27 HGTs and 57 loss events. A large number of incongruities are observed, most of them corresponding to transfers between species of the same taxonomic group and only two involving a transfer between groups. For both trees, GeneRax predicted HGTs in the deep branches to reconcile the order of appearance of the different subtrees. These incongruences are probably due to (i) the quality of the alignment as these sequences are poorly conserved resulting in an insufficient number of aligned sites containing a phylogenetic signal and (ii) methodological problems linked to the tree reconstruction approach and the substitution model that may not capture all the complexity of protein evolution even though the model used was determined as the best suited. For example, on the Hel308-Ski2 tree, the Methanomada and Thermoplasmata groups are not monophyletic as expected leading GeneRax to evoke HGTs to reconcile protein and species trees while this inconsistency may be due to insufficient signal in the multiple alignment. The same pertains for the Methanomada group in ASH-Ski2 tree. Regarding the number of loss events, GeneRax cannot infer the exact number of losses for at least one reason: if the gene is present in a species that is not sampled, GeneRax will interpret it as a loss. In conclusion, the GeneRax analysis of our data reveal that the genes encoding ASH-Ski2 and Hel308-Ski2 have evolved mainly by vertical inheritance, despite uncertainties about the deepest tree branches. Their absence in some genomes would most likely result from different independent gene loss events. However, the losses of ASH-Ski2 in Thermoplasmata and Archaeoglobales may have occurred only once in the common ancestor of each taxonomic group. In contrast, losses of Hel308-Ski2 are scattered in single genomes throughout the archaea taxonomy and may result from several independent events, except in Asgard where this could be due to incompleteness of their genome.

ASH-Ski2 and Hel308 of *Pyrococcus abyssi*

In order to gain a better understanding of the distinction between the biological functions of both groups, ASH-Ski2 and Hel308-Ski2 (denoted also as Hel308 in the following), we compared their biochemical activities using *Pyrococcus abyssi* as study model. Both proteins, ASH-Ski2 and Hel308, possess accessory domains in addition to the Ski2 core domain (Figure 3A). Hel308 possesses a Helix-Loop-Helix motif in its C-terminal region. ASH-Ski2 harbours a N-terminal extension with four strictly conserved cysteines that have the propensity to form in a zinc-finger domain (Supplementary Figure S2). Alignments of the predicted structures of ASH-Ski2$_{\Delta N}$ and Hel308 of *P. abyssi* obtained with AlphaFold software show a similar Ski2-core fold (RMSD = 1620) but they differ by three insertions for Hel308 and by an extra insertion of three α-helices in the ASH-Ski2$_{\Delta N}$ ratchet domain (Supplementary Figures S2 and S3). To evaluate the properties of ASH-Ski2 and Hel308, we purified wild-type and variants recombinant proteins (Supplementary Figure S4A). In this study, we used protein buffers with 300 mM of salt to stabilize *P. abyssi* proteins *in vitro*, as our previous studies have demonstrated (38). This salt concentration is also consistent with the intracellular ionic strength of approximately 350 mM determined for *P. furiosus* under physiological conditions (67). ASH-Ski2$_{\Delta N}$ and ASH-Ski2$_{K245A}$ recombinant proteins lack the specific ASH-Ski2 N-terminal domain and motif I known to be crucial for the binding of ATP, respectively. In addition, using gel exclusion chromatography, we showed that ASH-Ski2 is a monomeric enzyme (Supplementary Figure S4B). Accordingly, Hel308 was previously shown to be monomeric (35,36,68). Next, we compared the affinities of ASH-Ski2 and Hel308 to bind nucleic acids and their capacities to hydrolyse ATP, to unwind RNA and DNA overhang duplex substrates and to anneal RNA and DNA strands (Supplementary Table S3).

ASH-Ski2 and Hel308 nucleic acid binding affinities

To test the capacity of both proteins to bind single-stranded nucleic acids, we used a nitrocellulose-filter binding assay (Supplementary Figure S5A). Briefly, radiolabelled 50nt-DNA (DNA$_{50}$) or 50nt-RNA (RNA$_{50}$) substrates were incubated with increasing concentration of protein (0–200 nM). To determine the apparent dissociation constant (K_D), the percentage of bound fraction was plotted against protein concentration (Figure 3B). Similar K_D values were observed for Hel308, ASH-Ski2 and its variant proteins using DNA$_{50}$ or RNA$_{50}$ as substrates (Figure 3B and Supplementary Figure S5B). This indicates that ASH-Ski2 and Hel308 show no binding specificity either towards RNA or DNA molecule. Hence, it seems that the base specificity is not responsible for binding in nucleic acids. It is more likely that the phosphate backbone is involved in the binding process. Finally, the N-terminal extension of ASH-Ski2 does not hold a major nucleic acid binding site (Supplementary Figure S5B). Altogether, we showed that the presence of ATP did not affect nucleic acid affinity constants (Supplementary Figure S5C). Finally, we observed that the binding curves show a sigmoidal shape with a Hill coefficient superior to 1 (S-shape curves), indicating positive binding cooperativity or multiple binding sites. Most likely, more than one molecule of protein binds to multiple sites on a single molecule of 50 nt nucleic acids. This feature was also described in the case of the archaeal helicase aLhr2 (69). Further binding assays were conducted to compare the binding of single-stranded RNA (ssRNA) and double-stranded RNA (dsRNA) (Supplementary Figure S5D). The results suggest that ASH-Ski2 has a higher tendency to bind to ssRNA



Kd (nM)	ASH-Ski2			Hel308
	WT	ΔN	K245A	WT
RNA	20±4	29±2	29±4	39±4
DNA	24±3	17±3	34±3	32±4

Figure 3. (**A**) Domain protein organization of WT and variants of ASH-Ski2 and Hel308 from *P. abyssi*. Motif involved in ATP binding and hydrolysis are indicated in red, in coordination between nucleic acids and ATP binding in yellow and in nucleic acids binding in blue. ASH-Ski2$_{K245A}$ results from the mutation of the lysine 245 in alanine and ASH-Ski2$_{\Delta N}$ corresponds to the protein deleted for the N-terminal zinc finger domain. (**B**) Nucleic acid binding assays were performed using 0.5 nM of radiolabelled [γ-^{32}P]ATP RNA_{50} or DNA_{50} incubated with increasing concentrations of ASH-Ski2 (upper panel) and Hel308 (lower panel) ranging from 0 to 200 nM. RNA_{50} or DNA_{50}-protein complexes are retained on the nitrocellulose membrane while unbound RNA_{50} or DNA_{50} molecules are trapped on DEAE nylon membrane (Supplementary data, Figure S5A). The percentages of bound and unbound substrates are quantified over protein concentration to calculate the apparent affinity constant K_D values are indicated in the lower panel table. Values represent means of at least three independent experiments. (**C**) ATPase activity of ASH-Ski2 and Hel308 was measured by incubation of radiolabelled [γ-^{32}P]ATP at 65°C with 250nM of recombinant protein. Aliquots were taken after increasing times of incubation as indicated. The reaction products were analysed by TLC (Supplementary data Figure S6A). The extents of ATP hydrolysis are plotted over time. Each datum is an average of three separate experiments. Same experiments were carried out in presence of *P. abyssi* total RNA (+RNA). A control without protein is shown (Ø).

than dsRNA substrates, which is in line with the observation that Ski2-like helicases operate on overhang homoduplex substrates (7).

ASH-Ski2 and Hel308 are nucleic-acid dependent ATPases

To measure the capacity of ASH-Ski2 and Hel308 to hydrolyse ATP, we performed ATPase assays at 65°C with an excess of ATP, and in the absence or presence of total RNAs or DNAs extracted from *P. abyssi* cells. The release of inorganic phosphate was followed over time (Supplementary Figure S6A). Our results show that both total RNA and genomic DNA stimulate ASH-Ski2 and Hel308 ATPase activities (Figure 3C and Supplementary Figure S6B, respectively). In absence of nucleic acids, only residual ATP hydrolysis was observed. This is consistent with the general properties of helicases that adopt active states through conformational change mediated by nucleic acid binding (70). As observed for other SF2 helicases, ATP hydrolysis can be inactivated by mutating conserved motifs in the active site formed by the two RecA1 and RecA2 domains. The ASH-Ski2$_{K245A}$ protein mutated in the conserved motif I did not exhibit ATPase activity (Supplementary Figure S6C). On the other hand, the truncated ASH-Ski2$_{\Delta N}$, restricted to the Ski2-core domain, conserved its capacity to hydrolyse ATP meaning that the N-terminal extension domain of ASH-Ski2 does not interfere with ATP hydrolysis (Supplementary Figure S6C). We observed a noteworthy distinction in the ATP hydrolysis kinetics of ASH-Ski2 and Hel308, indicating a slower kinetics for ASH-Ski2 when compared to Hel308 (Figure 3C and Supplementary Figure S6B). It is therefore concluded that ASH-Ski2 has a lower level of ATPase activity than Hel308.

ASH-Ski2 and Hel308 have different unwinding and annealing activities and substrate specificities *in vitro*

Previous works reporting RNA or DNA helicase activities propose that ATP hydrolysis promotes the unwinding activity of helicase enzymes (8,71). In this regard, we predicted that ASH-Ski2 would have low unwinding activity compare to Hel308 given their respective ATPase activities reported above. To assess this hypothesis, we tested the capacity of both proteins to unwind nucleic acid duplexes by using 3′ or 5′ overhang RNA$_{50}$/RNA$_{26}$ or DNA$_{59}$/DNA$_{31}$ homoduplex substrates (Supplementary Table S3). These substrates were chosen on the basis of previous studies on *M. thermautotrophicus* and *P. abyssi* Hel308 (32,69). In our experimental setup, we used duplexes with the shortest strand labelled at its 5′ end. First, a protein-free control was done to assess temperature-dependent unwinding of RNA and DNA duplex substrates at 65°C (Figure 4A and Supplementary Figure S7A). We noticed, as expected, a weaker stability of DNA versus RNA duplexes which explain the increased lengths of used DNA substrates. The kinetics of strand dissociation were followed by incubating the pre-formed duplexes with purified proteins, in presence of ATP and of an excess of unlabelled oligo traps (Supplementary Figure S7A). This is to prevent the reverse pairing reaction by capturing the released single strand RNA$_{50}$ or DNA$_{59}$ generated over time. The percentage of newly-formed single strands was plotted over time. It took 40 min for ASH-Ski2 to unwind >75% of 3′ overhang 3′RNA$_{50}$/RNA$_{26}$ and <60% of 3′DNA$_{59}$/DNA$_{31}$ duplexes (Figure 4A, left panel). The unwinding activities were even less efficient (only 35%) for a 5′ overhang 5′RNA$_{50}$/RNA$_{26}$ duplex (Supplementary Figure S7B). Altogether, this indicates that ASH-Ski2 has a helicase activity with a preference for 3′ overhang RNA duplexes. The 3′ to 5′ unwinding polarity preference is consistent with the polarity observed for other members of the Ski2-like family including Hel308 from *M. thermautotrophicus* (7,32). Furthermore, ASH-Ski2$_{\Delta N}$ showed identical kinetics of strand dissociation as the wild-type protein indicating that the N-terminal domain is not determinant in this activity (Supplementary Figure S7C, upper panel). Surprisingly, a slower kinetics of strand dissociation was observed for ASH-Ski2$_{K245A}$ with no ATPase activity as shown above. This indicates that ATP hydrolysis is not a strict requirement for strand-dissociation activity, in our experimental conditions. This may explain why, in the absence of ATP, an activity of 50% is still observed for the wild-type protein (Supplementary Figure S7C, lower panel). It is possible that, in this case, the energy required for duplex separation is provided by the sole binding of the protein to the nucleic acid substrate. We performed similar assays with Hel308 which was reported to be a helicase with as strong DNA unwinding activity in *M. thermautotrophicus* and *S. solfataricus* (32,35,68). Indeed, and in contrast to ASH-Ski2, Hel308 could rapidly unwind 80% of 3′ overhang 3′DNA$_{59}$:DNA$_{31}$ duplexes after only 2 min of incubation but was unable to unwind 3′ overhang RNA duplexes (Figure 4A, right panel) as observed in previous work (69). Altogether, we concluded that ASH-Ski2 and Hel308 have significantly divergent activities and substrate preferences.

Furthermore, we tested the reverse reaction by investigating the ability of ASH-Ski2 and Hel308 in annealing RNA or DNA single strand oligonucleotides into homoduplexes by using the same set of oligonucleotides and experimental conditions (Figure 4B & Supplementary Figure S8). The percentage of newly formed duplexes was plotted overtime at 65°C. A control without protein was run in parallel to monitor spontaneous temperature dependent duplex formation (Supplementary Figure S8A). Remarkably, ASH-Ski2 assembled very efficiently 3′-overhang RNA homoduplexes (up to 80% in <2 min of incubation) (Figure 4B, left panel). It has been demonstrated that the level of protein concentration directly impacts this activity (Supplementary Figure S8B). In contrast, ASH-Ski2 assembled very poorly DNA substrates (Figure 4B, left panel and Supplementary Figure S8A). In this case, the percentage of formed duplexes is similar to that observed in the protein-free control. These results indicate that ASH-Ski2 has a robust strand-annealing activity with a clear preference for RNA substrates. This activity is not dependant of the N-terminal domain since identical kinetics were observed with the ASH-Ski2$_{\Delta N}$ protein variant (Supplementary Figure S8C). We conclude that, under our experimental conditions, ASH-Ski2 mainly adopts a state compatible with strand-annealing activity. Interestingly, ASH-Ski2$_{K245A}$ variant showed a reduced capacity of strand-annealing with a plateau reaching only 50% of duplexes indicating that only proportion of this protein variant adopts a state compatible with strand-annealing activity (Supplementary Figure S8C). In contrast to ASH-Ski2, Hel308 shows only a very low strand-pairing capacity, close to the control, regardless of the substrates used (Figure 4B, right panel). Altogether, we conclude that ASH-Ski2 and Hel308

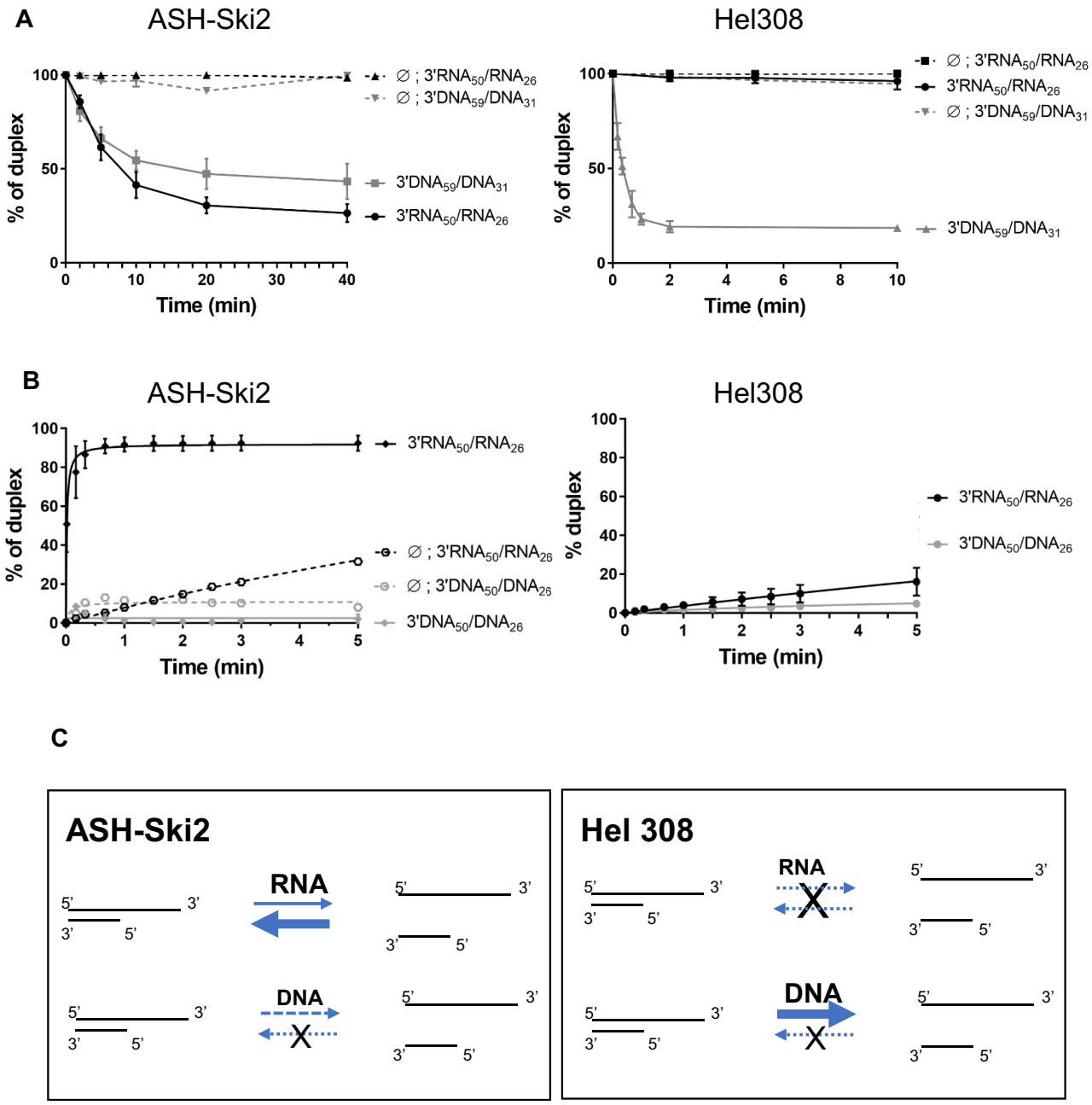


Figure 4. Unwinding and annealing activities of ASH-Ski2 and Hel 308. (**A**) Kinetics of strand dissociation by ASH-Ski2 (left panel) and Hel308 (right panel) in the presence of ATP are shown for 3′ overhang duplexes (3′RNA_{50}/RNA_{26} and 3′DNA_{59}/DNA_{31}). (**B**) Kinetics of strand association by ASH-Ski2 (left panel) and Hel308 (right panel) in absence of ATP are shown for 3′ overhang duplexes (3′RNA_{50}/RNA_{26} and 3′DNA_{59}/DNA_{31}). Protein free experiments were carried in parallel (Ø). Three independent experiments were performed for each condition. (**C**) Overview of the *in vitro* unwinding and annealing activities of *P. abyssi* ASH-Ski2 and Hel308 of this study. A cross indicates no detectable activity.

have significantly divergent *in vitro* activities and substrate preferences (Figure 4C).

Transcriptome analysis of *thermococcus barophilus* ASH-Ski2Δ strain

Considering the outlined activities of ASH-Ski2, we analysed its effect on RNA metabolism *in vivo* by conducting transcriptome analysis. As the present methods do not allow genome modification of *P. abyssi*, we opted to employ *T. barophilus*, a hyperthermophile Euryarchaeota species from the Thermococcales order that is closely related and phylogenetically similar, and was initially collected from the same environments. We were able to generate the ASH-Ski2Δ strain by deleting the TERMP_RS08745 gene (previously known as locus tag TERMP_01768) as detailed in reference (38). Considering that the ASH-Ski2 amino acid sequences of *P. abyssi* and *T. barophilus* share 90% similarity, it is postulated that the two orthologous proteins possess comparable biochemical properties. RNA was extracted from cultures that were exponentially grown and then subjected to RNA sequencing. The statistical analysis conducted showed that levels of 452 transcripts were significantly altered (with an adjusted *P*-value < 0.05). A total of 231 transcripts demonstrated

an increase in abundance, whereas the abundance of 221 transcripts decreased in the ΔASH-Ski2 strain, as opposed to the wild-type strain (Figure 5A and B; Supplementary Table S4). Amongst these differentially expressed genes, 60 demonstrated a change of more than two-fold, with 40 being up-regulated and 20 being down-regulated. To confirm the high throughput data, several genes were subjected to single-gene RT-qPCR (Supplementary Figure S9). The findings imply that ASH-Ski2 plays a role in the regulation of the level of 25% mRNAs. In addition, we observed a significant number of gene clusters, specifically 94 genes in 17 clusters, among the up-regulated transcripts. We defined gene clusters as a group of at least three neighbouring genes co-regulated and associated with the same activity. The largest group consisted of ribosomal proteins (TERMP_RS00440-00535), followed by genes involved in the histidine synthesis pathway (TERMP_RS02155-02200), genes encoding ABC-type transporters and membrane proteins (TERMP_RS02960-02975) and Na(+)-pumping/NADH-ubiquinone oxidoreductase ion transporters (TERMP_RS04255-04285). We carried out a GO enrichment term analysis to categorize impacted RNAs and to identify potentially affected cellular processes. Genes that were under-expressed in ΔASH-Ski2 strain did not show any significant GO enrichment. Two categories, namely, 'structural constituent of ribosome' (GO:0003735, MF) and 'ribosomal RNA binding proteins' (GO:0019843, MF), were significantly enriched among up-regulated genes due to an accumulation of transcripts encoding 24 ribosomal proteins, including 10 out of 25 proteins of the 30S subunit and 14 out of 39 proteins of the 50S subunit (Supplementary Table S5).

Discussion

Here, we investigated the phylogenetic diversity of archaeal Ski2-like family members and their biochemical activities using Thermococcales as models. The present study provides the first insight into the evolution of the Ski2-like helicase family in Archaea and the first evidence of Ski2-like members in Asgard genomes. We also discovered that, in total, three main groups of archaeal Ski2-like proteins exist, denoted as Hel308-Ski2, Brr2-Ski2 and ASH-Ski2, respectively. A phylogenetic divergence between ASH-Ski2 and Hel308 was observed and their respective activities were found to be different. Hel308 exhibits DNA helicase activity, initially documented in (30,31,32). In contrast, ASH-Ski2 demonstrates RNA helicase activity, as reported in here. The two different activities suggest different cellular roles.

Hel308-Ski2 DNA helicases are ubiquitous in Archaea

Our study shows that the major group of Ski2-like helicases that includes Hel308-Ski2 members is almost ubiquitous and should have evolved by vertical inheritance (Figure 1B and Supplementary Figure S1). We propose that the absence of Hel308-Ski2 in single genomes scattered throughout the archaea taxonomy results from independent loss events, except in Asgard where this may be due to the incompleteness of their genome. Eukaryotic PolQ/HelQ DNA helicases have been found to branch with the Hel308-Ski2 Asgard sequences (Figure 1B), consistent with the hypothesis that eukaryal domain emerged from Asgard archaea (1,2). Previous work assigned the function of archaeal Hel308 members to DNA metabolism (32,35,36). Consistently, by conducting series of biochemical assays to compare the *in vitro* enzymatic activities of both ASH-Ski2 and Hel308 proteins from the *P. abyssi* archaeon, our study confirms that Hel308 mainly acts on DNA substrates while ASH-Ski2 can act on both DNA and RNA. This discrepancy may be attributed to the existence of domain 5 (Helix Loop Helix) (termed ID-H in Supplementary Figures S2 and S3) in the C-terminal domain of Hel308, which serves as an autoinhibitory domain specifically towards DNA targets during unwinding activities (68). This region is not conserved in ASH-Ski2 (Supplementary Figures S2). The discrepancy within their C-terminal domains could account for the difference in substrate specificities between Hel308 and ASH-Ski2.

Unexpected discovery of the new group of Ski2-like helicases, Brr2-Ski2

Unexpectedly, this study also identifies a new subfamily of Ski2-like helicases, the archaeal Brr2-Ski2 group, which is restricted to Halobacteria and Methanocellales. Indeed, the archaeal Brr2-Ski2 clustered with Brr2 eukaryotic homologs under a same ancestral node, forming a well-supported subtree (100% SH-aLRT bootstrap support) which appears to emerge from the Hel308-Ski2 group (Figure 1B). Given tree location of archaeal sequences, narrow taxonomic distribution and the identification of the eukaryotic sec63 domain (Pfam PF02889) in the C-terminal part of some members, we propose that the archaeal Brr2-Ski2 sequences have been acquired through at least three independent horizontal gene transfer events (in the ancestor of Halobacteria, in the ancestor of two Methanocellales species, and in the Asgard *C. Heimdallarchaeota archaeon LC_3*) from eukaryotic genomes that are not sampled in our study (Figure 1B). To our knowledge, no study reports cellular functions of any halobacteria nor methanocellales Brr2 members. It was shown that the eukaryal Brr2 proteins are RNA helicases, with assigned roles in the dynamic regulation of the core splicing machinery (21,22,23). It should be noted that no pre-mRNA splicing like events involving a spliceosome like-machinery have been described in Archaea. Therefore, it is difficult to predict if archaeal Brr2-Ski2 members also act as RNA helicases in Halobacteria or Methanocellales.

ASH-Ski2 members are conserved RNA helicases in most Euryarchaea and Asgard

In addition, our study carried out with a large set of genomes, not only confirms our previous observation (38) showing that ASH-Ski2 group members are restricted to most Euryarchaeal group but also, unexpectedly, identified a new subset of ASH-Ski2 members in Asgard genomes (38 out of the 42 genomes considered in this study) (Figure 2). All ASH-Ski2 members contain a N-terminal domain with four strictly conserved cysteine residues, which have the propensity to form a zinc finger domain. Our analysis shows a good congruence between the archaeal species and the ASH-Ski2 trees. Therefore, we propose that genes encoding ASH-Ski2 were inherited vertically and that a member was encoded in the genome of the ancestor of the present-day Asgard, TACK and Euryarchaeota species. This gene would have been independently lost in the common ancestor of the three taxonomic groups, Archaeoglobales, Thermoplasmatales and TACK. In conclusion, this study settles that the ASH-Ski2 branch of the tree is monophyletic

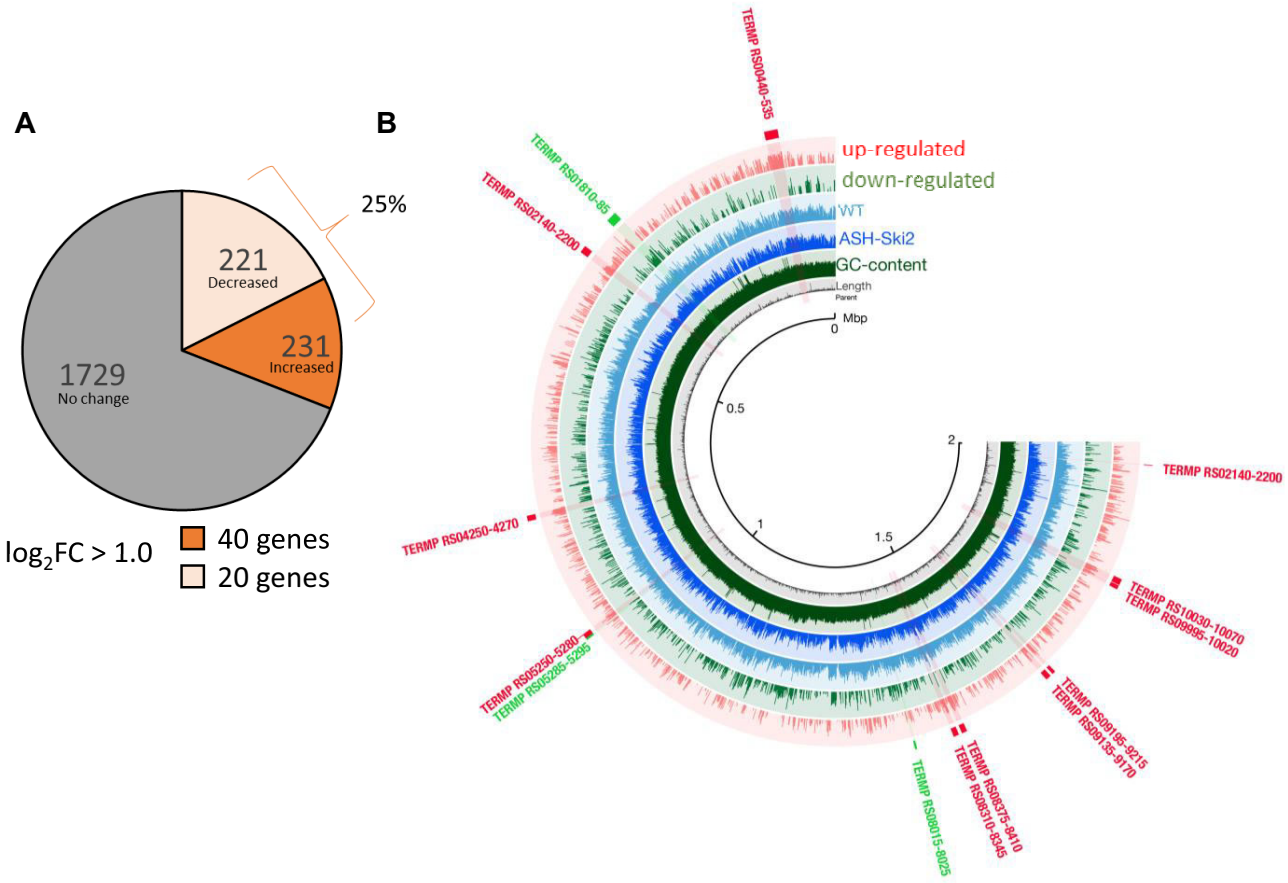


Figure 5. Effect of the deletion of the gene encoding ASH-Ski2 on the *T. barophilus* transcriptome. (**A**) 452 transcripts out of 2181 were differentially expressed in T. *barophilus* ASH-Ski2Δ strain. Decreased transcripts are shown in light orange and increased in dark orange. 60 transcripts showed >2-fold change (log_2FC higher than 1); 40 were up- and 20 down-regulated. (**B**) Read mapping and differential gene expression—the blue concentric circles represent the wild-type and ASH-Ski2Δ reads. Each ray represents a gene expression level. Red bars represent transcripts accumulated in ASH-Ski2Δ and green bars down-regulated transcripts, (ratio over wild-type). Some of the significantly up-regulated and down-regulated gene clusters are highlighted.

with only archaeal members and suggests an important physiological role of ASH-Ski2 in archaeal cells.

Our work also provides the first evidence that ASH-Ski2 from the *P. abyssi* archaeon supports an unwinding activity with a 3′-5′ polarity and, surprisingly, an efficient annealing activity to form RNA duplexes (summarized in Figure 4C). To assess the function of ASH-Ski2 RNA helicase in Thermococcales cells, we conducted a search for any effect of ASH-Ski2 on *T. barophilus* transcriptome landscape, but we found no clear pathway involving ASH-Ski2. We detected a deregulation of 25% of genes, with 3% being up- and down-regulated more than two-fold. Additionally, these changes in gene expression are not associated with a growth phenotype under optimal growth conditions. It is plausible that ASH-Ski2 plays a pivotal role in stress conditions that are yet to be identified. Moreover, certain deregulated genes are clustered on the genome and potentially form operons.

The individual functions of eukaryotic Mtr4 and Ski2 RNA helicases in nuclear and cytoplasmic RNA degradation are linked to specific nuclear and cytoplasmic ribonucleases, respectively. Additionally, the cytoplasmic Ski2, which is a cofactor of the SKI complex, has recently been demonstrated to play a crucial role in directing ribosome-bound RNA to the RNA exosome in humans (17). The eukaryotic Mtr4/Ski2 and ASH-Ski2 members are separated on the phylogenetic tree, it is conceivable that they exhibit divergent biochemical activities and perform distinct cellular functions. Notably, the eukaryotic Ski2 RNA helicase hydrolyses ATP to facilitate RNA unwinding in a processive manner with 3′-5′ polarity (7). The study demonstrates that ASH-Ski2 predominantly conducts annealing on RNA and exhibits minimal unwinding activity. It is also plausible that an unidentified factor stimulates the unwinding and ATPase activity, similar to eukaryal Mtr4 activated by ZCCHC8 (72). We tested whether aRNase J, which was shown to interact with ASH-Ski2 *in vitro* through its N-terminal domain, could affect ASH-Ski2 activity. We found no significant changes in either ATPase or annealing/unwinding activities. This is consistent with our finding that the N-terminal domain is not critical for ASH-Ski2 activity *in vitro* (Supplementary Figures S6C and S7C). However, other studies suggest that ASH-Ski2 might be involved in transcription termination (39,73). ASH-Ski2 from *Thermococcus kodakarensis* (denoted as Eta) has been shown to have a translocase activity *in vitro* (73). The interaction be-

tween ASH-Ski2/Eta and RNA polymerase is suggested to be mediated by the ratchet domain residues, rather than the N-terminal domain that is unique to the ASH-Ski2 group. In the reported studies, the ability of ASH-Ski2/Eta was only tested on DNA duplex substrates and not on RNA duplex substrates.

Concluding remarks

In conclusion, this work presents valuable evolutionary and functional insights into the Ski2-like helicase family, essential for further research in this field. We show that, in the course of evolution, ASH-Ski2 and Hel308-Ski2, the two main groups of archaeal Ski2-like proteins have diverged in structure and function. By interacting with a variety of adaptor proteins, these helicases may target different types of RNA or DNA substrates. Indeed, *S. tokodaii* and *S. islandicus* Hel308 were shown to interact with the Hjc endonuclease and the PINA ATPase, respectively, both involved in the resolution of Holliday junctions (35,36) while Thermococcale ASH-Ski2 was shown to interact with the 5′-3′ exoribonuclease, aRNase J (38). More specifically, this work opens many questions about the biological roles and structures of ASH-Ski2 RNA helicases identified in Euryarchaea and Asgard and on their spectre of action. It should be mentioned that aRNase J is absent from the Asgard phylum suggesting that ASH-Ski2 encoded in Asgard genomes may have a divergent function or it is possible that another unidentified ribonuclease may act together with ASH-Ski2 in Asgard archaea. Further characterisation of the molecular mechanisms coordinated by ASH-Ski2 protein helicases in these biological systems would enhance our understanding of their functions and evolution routes, notably in the context of the emergence of Eukarya from a branch within the Asgard phylum (1,2). Next challenges are to understand how ASH-Ski2 helicases act, recognize their RNA targets and to determine if the link between ASH-Ski2 helicases and the translation apparatus exist in the context of RNA surveillance pathways as observed in Eukarya. Finally, the coordination of the annealing/unwinding activities of ASH-Ski2 with the actions of RNA-degrading machines remains to be understood. Interestingly, a variety of proteins with biological roles ranging from bacterial translational regulation to the maturation of eukaryotic transcripts, RNA editing and viral infection have been shown to accelerate RNA/RNA annealing (74,75,76,77,78).

Data availability

The Transcriptomic data mass have been deposited via the GEO repository with the dataset identifier GSE229955 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE229955).

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

We are indebted to Y. Quentin for his expertise in taxonomic identification of archaeal helicase members, M. Aguirrebengoa for her expertise in RNA seq analysis (BigA service), M. Georges for technical help on RNA preparation, M. Hajj for technical help on activity assays of Hel308 and R. Dulermo for genetic screens and helpful discussions.

Funding

CNRS, University Paul Sabatier, French Ministère de l'Enseignement Supérieure et de la Recherche and French 'Agence Nationale pour la Recherche' [ANR-16-CE12-0016-01 to B.C.O., ANR-22-CE12-0009-01 to B.C.O.]. Funding for open access charge: Agence Nationale pour la Recherche.

Conflict of interest statement

None declared.

References

1. Nobs,S.-J., MacLeod,F.I., Wong,H.L. and Burns,B.P. (2022) Eukarya the chimera: eukaryotes, a secondary innovation of the two domains of life?*Trends Microbiol.*, **30**, 421–431.
2. Spang,A., Saw,J.H., Jørgensen,S.L., Zaremba-Niedzwiedzka,K., Martijn,J., Lind,A.E., van Eijk,R., Schleper,C., Guy,L. and Ettema,T.J.G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, **521**, 173–179.
3. Spang,A., Caceres,E.F. and Ettema,T.J.G. (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science*, **357**, 6351.
4. Eme,L., Spang,A., Lombard,J., Stairs,C.W. and Ettema,T.J.G. (2017) Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.*, **15**, 711–723.
5. Tahon,G., Geesink,P. and Ettema,T.J.G. (2021) Expanding archaeal diversity and phylogeny: past, present, and future. *Annu. Rev. Microbiol.*, **75**, 359–381.
6. Chamieh,H., Ibrahim,H. and Kozah,J. (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene*, **576**, 214–228.
7. Johnson,S.J. and Jackson,R.N. (2013) Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol*, **10**, 33–43.
8. Fairman-Williams,M.E., Guenther,U.-P. and Jankowsky,E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
9. Büttner,K., Nehring,S. and Hopfner,K.-P. (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.*, **14**, 647–652.
10. Falk,S., Tants,J.-N., Basquin,J., Thoms,M., Hurt,E., Sattler,M. and Conti,E. (2017) Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53. *RNA*, **23**, 1780–1787.
11. Weick,E.-M., Puno,M.R., Januszyk,K., Zinder,J.C., DiMattia,M.A. and Lima,C.D. (2018) Helicase-dependent RNA decay illuminated by a cryo-EM structure of a Human nuclear RNA exosome-MTR4 complex. *Cell*, **173**, 1663–1677.
12. Halbach,F., Rode,M. and Conti,E. (2012) The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA*, **18**, 124–134.
13. Weick,E.-M. and Lima,C.D. (2021) RNA helicases are hubs that orchestrate exosome-dependent 3′-5′ decay. *Curr. Opin. Struct. Biol.*, **67**, 86–94.
14. Patrick,E.M., Srinivasan,S., Jankowsky,E. and Comstock,M.J. (2017) The RNA helicase Mtr4p is a duplex-sensing translocase. *Nat. Chem. Biol.*, **13**, 99–104.
15. Bernstein,J., Patterson,D.N., Wilson,G.M. and Toth,E.A. (2008) Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3′→5′ helicase partner of the nuclear exosome. *J. Biol. Chem.*, **283**, 4930–4942.

16. Meola,N., Domanski,M., Karadoulama,E., Chen,Y., Gentil,C., Pultz,D., Vitting-Seerup,K., Lykke-Andersen,S., Andersen,J.S., Sandelin,A., *et al.* (2016) Identification of a nuclear exosome decay pathway for processed transcripts. *Mol. Cell*, **64**, 520–533.
17. Kögel,A., Keidel,A., Bonneau,F., Schäfer,I.B. and Conti,E. (2022) The human SKI complex regulates channeling of ribosome-bound RNA to the exosome via an intrinsic gatekeeping mechanism. *Mol. Cell*, **82**, 756–769.
18. Halbach,F., Reichelt,P., Rode,M. and Conti,E. (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*, **154**, 814–826.
19. Schmidt,C., Kowalinski,E., Shanmuganathan,V., Defenouillère,Q., Braunger,K., Heuer,A., Pech,M., Namane,A., Berninghausen,O., Fromont-Racine,M., *et al.* (2016) The cryo-EM structure of a ribosome-Ski2-Ski3-Ski8 helicase complex. *Science*, **354**, 1431–1433.
20. Zinoviev,A., Ayupov,R.K., Abaeva,I.S., Hellen,C.U.T. and Pestova,T.V. (2020) Extraction of mRNA from stalled ribosomes by the ski complex. *Mol. Cell*, **77**, 1340–1349.
21. Lauber,J., Fabrizio,P., Teigelkamp,S., Lane,W.S., Hartmann,E. and Luhrmann,R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. *EMBO J.*, **15**, 4001–4015.
22. Zhang,L., Li,X., Hill,R.C., Qiu,Y., Zhang,W., Hansen,K.C. and Zhao,R. (2015) Brr2 plays a role in spliceosomal activation in addition to U4/U6 unwinding. *Nucleic Acids Res.*, **43**, 3286–3297.
23. Vester,K., Santos,K.F., Kuropka,B., Weise,C. and Wahl,M.C. (2020) The inactive C-terminal cassette of the dual-cassette RNA helicase BRR2 both stimulates and inhibits the activity of the N-terminal helicase unit. *J. Biol. Chem.*, **295**, 2097–2112.
24. Sitron,C.S., Park,J.H. and Brandman,O. (2017) Asc1, Hel2, and Slh1 couple translation arrest to nascent chain degradation. *RNA*, **23**, 798–810.
25. Matsuo,Y., Ikeuchi,K., Saeki,Y., Iwasaki,S., Schmidt,C., Udagawa,T., Sato,F., Tsuchiya,H., Becker,T., Tanaka,K., *et al.* (2017) Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat. Commun.*, **8**, 159.
26. Best,K., Ikeuchi,K., Kater,L., Best,D., Musial,J., Matsuo,Y., Berninghausen,O., Becker,T., Inada,T. and Beckmann,R. (2023) Structural basis for clearing of ribosome collisions by the RQT complex. *Nat. Commun.*, **14**, 921.
27. Tanaka,K., Miyamoto,N., Shouguchi-Miyata,J. and Ikeda,J.-E. (2006) HFM1, the human homologue of yeast Mer3, encodes a putative DNA helicase expressed specifically in germ-line cells. *DNA Seq.*, **17**, 242–246.
28. Guiraldelli,M.F., Eyster,C., Wilkerson,J.L., Dresser,M.E. and Pezza,R.J. (2013) Mouse HFM1/Mer3 is required for crossover formation and complete synapsis of homologous chromosomes during meiosis. *PLoS Genet.*, **9**, e1003383.
29. Wood,R.D. and Doublié,S. (2016) DNA polymerase θ (POLQ), double-strand break repair, and cancer. *DNA Repair (Amst.)*, **44**, 22–32.
30. Fujikane,R., Komori,K., Shinagawa,H. and Ishino,Y. (2005) Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus. *J. Biol. Chem.*, **280**, 12351–12358.
31. Fujikane,R., Shinagawa,H. and Ishino,Y. (2006) The archaeal hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork. *Genes Cells*, **11**, 99–110.
32. Guy,C.P. and Bolt,E.L. (2005) Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res.*, **33**, 3678–3690.
33. Marini,D.B., Zhang,Y.P., Rowhani,A. and Uyemoto,J.K. (2002) Etiology and host range of a closterovirus associated with plum bark necrosis-stem pitting disease. *Plant Dis.*, **86**, 415–417.
34. Tafel,A.A., Wu,L. and McHugh,P.J. (2011) Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures. *J. Biol. Chem.*, **286**, 15832–15840.
35. Li,Z., Lu,S., Hou,G., Ma,X., Sheng,D., Ni,J. and Shen,Y. (2008) Hjm/Hel308A DNA helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with hjc endonuclease in vitro. *J. Bacteriol.*, **190**, 3006–3017.
36. Zhai,B., DuPrez,K., Han,X., Yuan,Z., Ahmad,S., Xu,C., Gu,L., Ni,J., Fan,L. and Shen,Y. (2018) The archaeal ATPase PINA interacts with the helicase hjm via its carboxyl terminal KH domain remodeling and processing replication fork and Holliday junction. *Nucleic Acids Res.*, **46**, 6627–6641.
37. Lever,R.J., Simmons,E., Gamble-Milner,R., Buckley,R.J., Harrison,C., Parkes,A.J., Mitchell,L., Gausden,J.A., Škulj,S., Bertoša,B., *et al.* (2023) Archaeal Hel308 suppresses recombination through a catalytic switch that controls DNA annealing. *Nucleic Acids Res.*, **51**, 8563–8574.
38. Phung,D.K., Etienne,C., Batista,M., Langendijk-Genevaux,P., Moalic,Y., Laurent,S., Liuu,S., Morales,V., Jebbar,M., Fichant,G., *et al.* (2020) RNA processing machineries in Archaea: the 5′-3′ exoribonuclease aRNase J of the β-CASP family is engaged specifically with the helicase ASH-Ski2 and the 3′-5′ exoribonucleolytic RNA exosome machinery. *Nucleic Acids Res.*, **48**, 3832–3847.
39. Walker,J.E., Luyties,O. and Santangelo,T.J. (2017) Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E6767–E6773.
40. Maier,L.-K. and Marchfelder,A. (2019) It's all about the T: transcription termination in archaea. *Biochem. Soc. Trans.*, **47**, 461–468.
41. Lu,S., Wang,J., Chitsaz,F., Derbyshire,M.K., Geer,R.C., Gonzales,N.R., Gwadz,M., Hurwitz,D.I., Marchler,G.H., Song,J.S., *et al.* (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.*, **48**, D265–D268.
42. Kim,D.H. and Rossi,J.J. (1999) The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA*, **5**, 959–971.
43. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
44. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
45. Ali,R.H., Bogusz,M. and Whelan,S. (2019) Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol. Biol. Evol.*, **36**, 2340–2351.
46. Darriba,D., Posada,D., Kozlov,A.M., Stamatakis,A., Morel,B. and Flouri,T. (2020) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.*, **37**, 291–294.
47. Minh,B.Q., Schmidt,H.A., Chernomor,O., Schrempf,D., Woodhams,M.D., Haeseler,A. and Lanfear,R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, **37**, 1530–1534.
48. Letunic,I. and Bork,P. (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49**, W293–W296.
49. Parks,D.H., Rinke,C., Chuvochina,M., Chaumeil,P.-A., Woodcroft,B.J., Evans,P.N., Hugenholtz,P. and Tyson,G.W. (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.*, **2**, 1533–1542.
50. Eddy,S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.*, **7**, e1002195.
51. Capella-Gutiérrez,S., Silla-Martínez,J.M. and Gabaldón,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
52. Chang,J.-M., Di Tommaso,P. and Notredame,C. (2014) TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.*, **31**, 1625–1637.

53. Williams,T.A., Szöllősi,G.J., Spang,A., Foster,P.G., Heaps,S.E., Boussau,B., Ettema,T.J.G. and Embley,T.M. (2017) Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E4602–E4611.
54. Morel,B., Kozlov,A.M., Stamatakis,A. and Szöllősi,G.J. (2020) GeneRax: a tool for species-tree-aware maximum likelihood-based gene Family tree inference under gene duplication, transfer, and loss. *Mol. Biol. Evol.*, **37**, 2763–2774.
55. Phok,K., Moisan,A., Rinaldi,D., Brucato,N., Carpousis,A.J., Gaspin,C. and Clouet-d'Orval,B. (2011) Identification of CRISPR and riboswitch related RNAs among novel noncoding RNAs of the euryarchaeon Pyrococcus abyssi. *Bmc Genomics (Electronic Resource)*, **12**, 312.
56. Thiel,A., Michoud,G., Moalic,Y., Flament,D. and Jebbar,M. (2014) Genetic manipulations of the hyperthermophilic piezophilic archaeon thermococcus barophilus. *Appl. Environ. Microbiol.*, **80**, 2299–2306.
57. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
58. Putri,G.H., Anders,S., Pyl,P.T., Pimanda,J.E. and Zanini,F. (2022) Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics*, **38**, 2943–2945.
59. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
60. Carlson,M. and Pagès,H. (2019) AnnotationForge: tools for building SQLite-based annotation data packages. R package version.
61. Wu,T., Hu,E., Xu,S., Chen,M., Guo,P., Dai,Z., Feng,T., Zhou,L., Tang,W., Zhan,L., *et al.* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)*, **2**, 100141.
62. Eren,A.M., Esen,Ö.C., Quince,C., Vineis,J.H., Morrison,H.G., Sogin,M.L. and Delmont,T.O. (2015) Anvi'o: an advanced analysis and visualization platform for' ' 'omics data. *PeerJ*, **3**, e1319.
63. Moalic,Y., Hartunians,J., Dalmasso,C., Courtine,D., Georges,M., Oger,P., Shao,Z., Jebbar,M. and Alain,K. (2021) The piezo-hyperthermophilic archaeon thermococcus piezophilus regulates its energy efficiency system to cope with large hydrostatic pressure variations. *Front. Microbiol.*, **12**, 730231.
64. Pyle,A.M. (2011) RNA helicases and remodeling proteins. *Curr. Opin. Chem. Biol.*, **15**, 636–642.
65. Weir,J.R., Bonneau,F., Hentschel,J. and Conti,E. (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 12139–12144.
66. Eme,L. and Ettema,T.J.G. (2018) The eukaryotic ancestor shapes up. *Nature*, **562**, 352–353.
67. Hasan,M.N., Hagedoorn,P.L. and Hagen,W.R. (2002) Pyrococcus furiosus ferredoxin is a functional dimer. *FEBS Lett.*, **531**, 335–338.
68. Richards,J.D., Johnson,K.A., Liu,H., McRobbie,A.-M., McMahon,S., Oke,M., Carter,L., Naismith,J.H. and White,M.F. (2008) Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *J. Biol. Chem.*, **283**, 5118–5126.
69. Hajj,M., Langendijk-Genevaux,P., Batista,M., Quentin,Y., Laurent,S., Capeyrou,R., Abdel-Razzak,Z., Flament,D., Chamieh,H., Fichant,G., *et al.* (2021) Phylogenetic diversity of lhr proteins and biochemical activities of the thermococcales aLhr2 DNA/RNA helicase. *Biomolecules*, **11**, 950.
70. Jankowsky,E. and Fairman,M.E. (2007) RNA helicases–one fold for many functions. *Curr. Opin. Struct. Biol.*, **17**, 316–324.
71. Pyle,A.M. (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.*, **37**, 317–336.
72. Puno,M.R. and Lima,C.D. (2018) Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E5506–E5515.
73. Marshall,C.J., Qayyum,M.Z., Walker,J.E., Murakami,K.S. and Santangelo,T.J. (2022) The structure and activities of the archaeal transcription termination factor eta detail vulnerabilities of the transcription elongation complex. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2207581119.
74. Rajkowitsch,L., Chen,D., Stampfl,S., Semrad,K., Waldsich,C., Mayer,O., Jantsch,M.F., Konrat,R., Bläsi,U. and Schroeder,R. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol*, **4**, 118–130.
75. Doetsch,M., Fürtig,B., Gstrein,T., Stampfl,S. and Schroeder,R. (2011) The RNA annealing mechanism of the HIV-1 Tat peptide: conversion of the RNA into an annealing-competent conformation. *Nucleic Acids Res.*, **39**, 4405–4418.
76. Portman,D.S. and Dreyfuss,G. (1994) RNA annealing activities in HeLa nuclei. *EMBO J.*, **13**, 213–221.
77. Müller,U.F., Lambert,L. and Göringer,H.U. (2001) Annealing of RNA editing substrates facilitated by guide RNA-binding protein gBP21. *EMBO J.*, **20**, 1394–1404.
78. Kwon,S.-H., Lee,I.-H., Kim,N.-Y., Choi,D.-H., Oh,Y.-M. and Bae,S.-H. (2007) Translation initiation factor eIF1A possesses RNA annealing activity in its oligonucleotide-binding fold. *Biochem. Biophys. Res. Commun.*, **361**, 681–686.

Received: May 10, 2023. **Revised:** February 19, 2024. **Editorial Decision:** February 26, 2024. **Accepted:** February 27, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com