nature communications

Article

https://doi.org/10.1038/s41467-024-50195-2

Asgard archaea defense systems and their roles in the origin of eukaryotic immunity

Received: 16 April 2024

Accepted: 28 June 2024

Published online: 31 July 2024

Check for updates

Pedro Leão[1,2,5] ✉, Mary E. Little[3], Kathryn E. Appler[2], Daphne Sahaya[3], Emily Aguilar-Pine[1], Kathryn Currie[1], Ilya J. Finkelstein[3,4], Valerie De Anda[1,2] & Brett J. Baker[1,2] ✉

Dozens of new antiviral systems have been recently characterized in bacteria. Some of these systems are present in eukaryotes and appear to have originated in prokaryotes, but little is known about these defense mechanisms in archaea. Here, we explore the diversity and distribution of defense systems in archaea and identify 2610 complete systems in *Asgardarchaeota*, a group of archaea related to eukaryotes. The Asgard defense systems comprise 89 unique systems, including argonaute, NLR, Mokosh, viperin, Lassamu, and CBASS. Asgard viperin and argonaute proteins have structural homology to eukaryotic proteins, and phylogenetic analyses suggest that eukaryotic viperin proteins were derived from Asgard viperins. We show that Asgard viperins display anti-phage activity when heterologously expressed in bacteria. Eukaryotic and bacterial argonaute proteins appear to have originated in *Asgardarchaeota*, and Asgard argonaute proteins have argonaute-PIWI domains, key components of eukaryotic RNA interference systems. Our results support that *Asgardarchaeota* played important roles in the origin of antiviral defense systems in eukaryotes.

Organisms across the tree of life contain complex defense systems (DS) to battle viral infections[1–3]. Over the past decade, dozens of new DS have been identified and characterized in bacteria, sparking a debate about a potential link between these systems and the origins of innate immune mechanisms in eukaryotes. More recently, protein components of bacterial NLR (Nucleotide-binding domain leucine-rich repeat), CBASS (Cyclic oligonucleotide-based antiphage signaling system), viperins (virus-inhibitory protein, endoplasmic reticulum-associated, interferon (IFN)-inducible), argonautes, and other DS have been shown to exhibit homology with proteins involved in the eukaryotic immune system[4]. Most of the research on prokaryotic defense systems has focused on bacteria, with archaea representing <3% of the genomes in these studies[5–7]. Thus, very little is known about the diversity or evolution of these systems in archaea.

Recently, diverse novel genomes have been obtained belonging to the archaea most closely related to eukaryotes, commonly referred to as "Asgard" archaea, the phylum Asgardarchaeota[8]. In addition to being sister lineages to eukaryotes, these archaea also contain an array of genes that are hallmarks of complex cellular life involved in signal processing, transcription, and translocations, among other processes[9]. The Asgard archaea are descendants of the ancestral host that gave rise to eukaryotic life. One newly described order, the Hodarchaeales (within the Heimdallarchaeia class), shared a common ancestor with eukaryotes[8]. Here, we characterize defense systems in archaea and show that Asgard archaea have a broad array of these DS. We also show that Asgards contributed to the origins of innate immune mechanisms in eukaryotes.

[1]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA. [2]Department of Marine Science, Marine Science Institute, University of Texas at Austin, Port Aransas, TX, USA. [3]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA. [4]Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX, USA. [5]Present address: Department of Microbiology - RIBES, Radboud University, Nijmegen, The Netherlands. ✉e-mail: pedro.leao@ru.nl; acidophile@gmail.com

https://doi.org/10.1038/s41467-024-50195-2

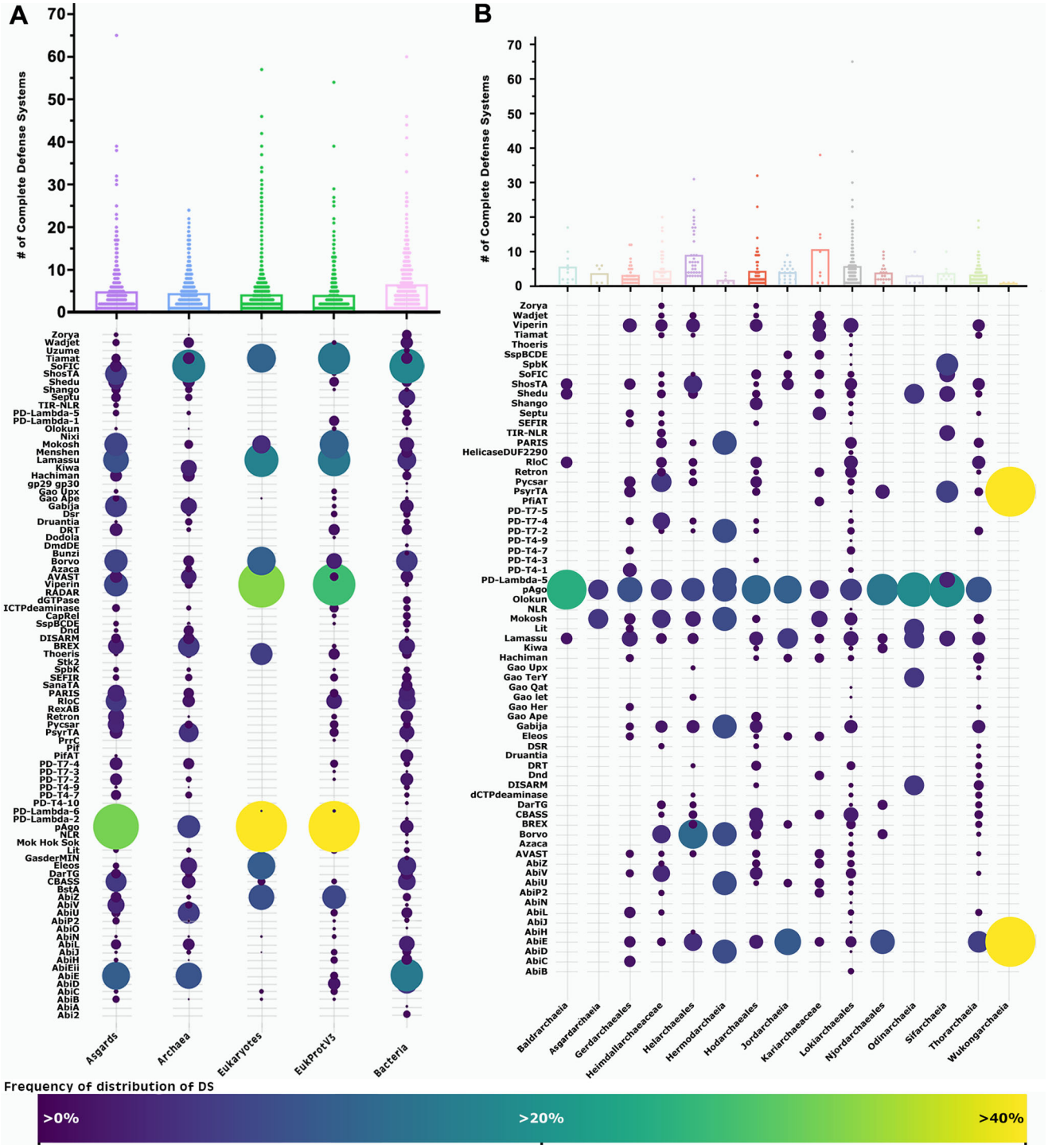

Fig. 1 | Distribution of defense systems across the tree of life and in Asgard archaea. A The total number of complete defense systems (DS) in each genomic dataset used in this study. The bubble plot shows the frequency of distribution of DS identified (Y-axis) in the datasets (X-axis). **B** The total number of DS across Asgardarchaeota genomes. Bubble plot: frequency of distribution of DS in each Asgardarchaeota group. Dots in the graphs represent the number of DS in a single genome from the respective group. The size and the color of the dots in the bubble plot are proportional to the prevalence of the DS in that group. For clarity, the 2 most prevalent DS (CRISPR and RM) across the genomic datasets were removed from both bubble plots representation. DS less found in Eukaryotes (check Methods section for the list) were also removed from the bubble plot in panel A.

Results and discussion

Distribution of Defense Systems across the tree of life

To explore the diversity and distribution of DS in archaea, we used 132 previously described defense systems to search through a comprehensive dataset containing 3408 publicly available genomes, and a newly expanded set of Asgardarchaeota genomes (869). A total of 30,761 defense system-associated genes have been identified across these archaeal genomes, belonging to 11,466 complete defense systems, with an average of 4.55 DS genes per archaeal genome (Fig. 1; Supplementary Data 1). Of a total of 132 previously described DS, 89 were found in these genomes. 27.5% of these archaeal genomes do not contain any known DS. A total of 2610 complete DS were identified in Asgardarchaeota genomes. These varied among the classes between 1 and 65, with an average of 4.9 DS per genome. These numbers are

similar to those reported in bacteria, on average 5.6 DS identified by Tesson et al.[6] and 5.8 DS by Millman et al.[7]. The DPANN archaea contain the lowest number, with a mean of 2.6 DS per genome. As previously reported in bacteria, the most prevalent DS in archaea are the restriction-modification (RM) and CRISPR systems[6,10]. These DSs represent 45% and 22% of all known defense systems in archaeal genomes, respectively (Supplementary Data 1).

Asgardarchaeota have a broad array of DS compared to those in bacteria. Some Asgard archaeal groups have a higher DS per genome ratio compared to other archaea (Baldrarchaeia 5.7), and others surpass bacteria (Helarchaeales 9.1; Kariarchaeaceae 10.7; Lokiarchaeales 5.8). Four out of the five representatives from the Heimdallarchaeia class, including genomes from Hodarchaeales, have fewer DS (Njordarchaeales 4.0; Gerdarchaeales 3.3; Heimdallarchaeaceae 4.5; Hodarchaeales 4.4) (Fig. 1B; Graph), suggesting they have yet uncharacterized DS. Given that Heimdallarchaeia is the prokaryotic class most closely related to eukaryotes[8], it is possible that these organisms might utilize mechanisms more similar to those observed in unicellular eukaryotes that were not detected using a prokaryotic DS database.

Twenty-two DS are more frequently found in Asgardarchaeota genomes compared to other prokaryotes (AbiB, AbiP2, AbiV, Argonautes, Borvo, Cas, dCTPdeaminase, Gao Ape, Gao Her, Gao Iet, hachiman, pycsar, PD-T7-4, PD-T4-1, PD-T4-7, RloC, Rst helicase, Rst2TM TIR-NLR, ShosTA, Shango, Shedu, viperins). Conversely, 12 DS are more abundant in archaea from groups outside of Asgardarchaeota (Fig. 2B; Bubble plot). Viperins and argonautes account for 2.1% and 7.9% (respectively) of all DS identified in Asgardarchaeota genomes. Viperins and argonautes are found 9 and 4 times more frequently in Asgard genomes than in other archaeal groups. When compared to bacteria, the disparity is even more pronounced, with 29 and 11 times higher representation of viperins and argonautes in Asgardarchaeota. Argonautes are present in every Asgardarchaeota class.

Asgard archaea and Eukaryotic viperins shared a common origin

Viperins were first described in eukaryotes as one of the key players in the mechanism of inhibition of human cytomegalovirus (HCMV) infection[11] and later found in prokaryotes[3]. Viperin proteins in archaea and bacteria have sequence and structural residues conservation[12]. These structural residues are a strong indicator of a conserved defense mechanism[13]. Protein sequence and structural conservations make it possible to reconstruct protein phylogenies, which indicate a prokaryotic origin and show Asgard archaea as being a sister group to eukaryotes, suggesting eukaryotic and Asgard viperins evolved from a common protein ancestor (Fig. 2).

Using an expanded archaeal genomic dataset our phylogenies revealed eukaryotic viperins (eVip) and Asgard viperins (asVip) are sister proteins and share a common ancestor (Fig. 2A). Only four proteins outside the Asgard group (one archaeal and three bacterial) are present in this clade. The two bacterial species harboring these viperin sequences are cyanobacteria belonging to the genera *Anabaena* and *Planktothricoides*, known for their symbiotic relationships with eukaryotic organisms. *Anabaena* spp. are present throughout the entire lifecycle of plants from the *Azolla* genus[14]. Co-speciation and gene transfer have been reported in these host-symbiont interactions[15], which could explain the position of these viperins near eukaryotes. The diversity of basal asVip to eukaryotic sequences suggests that this immune mechanism in eukaryotes was derived from Asgardarchaeota (Fig. 2A, red nodes).

Viperin proteins are structurally conserved across all domains of life (Fig. 2B). As a member of the radical S-adenosylmethionine superfamily, the SAM domain is present in all viperins characterized by a partial (βα)6-barrel folds located at the center of the protein[16]. The most conserved structure is the catalytic site cavity closer to the C-terminal extension. This site is present in asVips (yellow in Fig. 2C), indicating that the 3′-deoxy-3′,4′-didehydro (ddh) synthase activity has been preserved among all domains of life (Fig. 2C). Concomitant with our study, Shomar et al. provided experimental support for this proposed conservation[17]. These findings provide evidence for its conserved function as a defense mechanism, since ddh is the primary component of this DS, acting as the chain terminator in viral DNA/RNA polymerase in eukaryotes[18], and prokaryotes[13], and potentially in Asgardarchaeota (Fig. 2D, E; discuss below).

To ensure that asVips can protect cells against viral infections, we challenged asVip-expressing bacteria with T7 phage. Both bacterial and eVips have previously been demonstrated to inhibit this phage's infection[13]. We synthesized and constitutively expressed 48 Asgard viperins in *E. coli*, which were then submitted to testing (Supplementary Data 2). At first, 9 out of the 48 asVips (19%), all belonging to Lokiarchaeales group, displayed anti-T7 defense activity (Fig. 2D).

In a second assay, *asVip* were redesigned with codon optimization for heterologous expression in *E. coli*. Remarkably, a viperin originating from a Hodarchaeales genome was then capable of safeguarding cells against T7 phage infection (Fig. 2E). Asgard archaea are known for possessing ribosomal architecture akin to eukaryotes[19]. More recently, representatives from Hodarchaeales have been found to contain a homologue of the previously eukaryote-exclusive ribosomal protein L28e[8]. The lack of viral protection observed in our initial assay could be indicative of the inefficiency of bacterial systems in expressing Asgard viperins from groups more closely related to eukaryotes, such as Hodarchaeales. This obstacle was overcome following codon optimization, which allowed the protein to be expressed more efficiently and demonstrate its effectiveness in protecting cells against viral infection. These analyses show that asVips are consistent with the defensive roles against viral infection observed for viperins previously described in both bacterial and eukaryotic contexts[13,18].

Cross-Domain conservation of sequence and structure of argonautes

In eukaryotes, both microRNAs (miRNAs) and small interfering RNAs (siRNAs) serve as guides that target RNA transcripts for degradation by the RNA-induced silencing complex (RISC) as a form of post-transcriptional regulation and cellular defense[20]. The signature proteins of this complex are the argonautes[21]. These proteins are members of the PIWI (P element–induced wimpy testis) superfamily, distinguished by the presence of the homonyms domain[22]. Argonaute-PIWI proteins are key components of the RNA interference (RNAi) system in eukaryotes. Prokaryotic argonautes (pAgos) are also associated with defense against mobile genetic elements[6,23–25]. pAgos are classified into long pAgos and short pAgos. Long pAgos domains organization is similar to that of eukaryotic argonautes (eAgos), comprising a N-terminal domain that serves as a wedge to separate the guide oligonucleotide from its target, a MID (middle) domain, and a PAZ (PIWI-Argonaute-Zwille) domain that together, anchors and stabilizes the guide molecule. Long pAgos also have a catalytic PIWI domain responsible for the cleavage of the target DNA/RNA[26,27]. Short pAgos possess only the MID and an inactive PIWI domain[24].

Previously, the origin of eAgo has been proposed to be from euryarchaeal argonautes, and an ancient phylogenetic split between short and long types of these proteins was observed[23,24,28]. Phylogenetic analysis of archaeal argonautes (arAgo) identified in this study initially supported the early bifurcation into short and long forms (Supplementary Fig. 1). However, it also revealed that Asgardarchaeota argonautes (asAgo) are the likely precursor of eAgo (Fig. 3A, red nodes; Supplementary Fig. 1). Our phylogenetic analysis also reveals a deep-rooted clade of asAgo as the ancestral form of long bacterial and eukaryotes argonautes (Fig. 3A). An early clade of asAgo diversified into a larger group that encompasses arAgo and more than 98% of eAgo (Fig. 3A). Interestingly, we also found that each clade of long proteins across the phylogeny is rooted by asAgos to some extent. This

https://doi.org/10.1038/s41467-024-50195-2

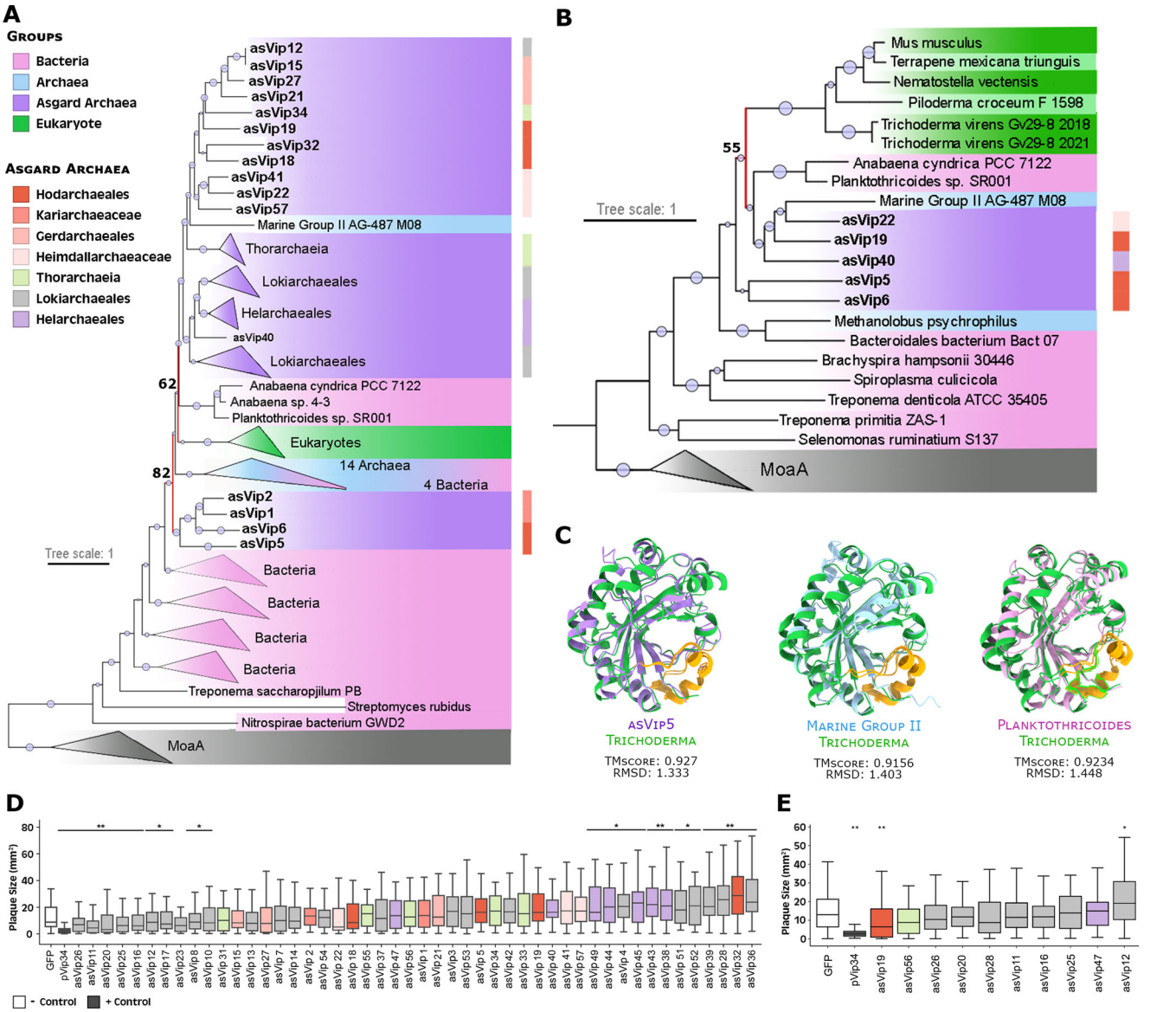

Fig. 2 | Evolutionary history and anti-phage activity of Asgard viperins. **A** Phylogenetic analysis of viperins. Viperins phylogeny revealed ancestral links of eVip (eukaryotic viperin) with asVip (asgard viperin) (nodes marked in red), particularly those within the Heimdallarchaeia class (including Kariarchaeaceae (2), Heimdallarchaeaceae (3) and Hodarchaeales (5)). The size of the dots on the nodes is proportional to bootstrap values ranging between 60 and 100. **B** Structure-based homology of viperins. Consistent with the sequence homology-based phylogenetic tree, the eVip structure appears to have been inherited from asVip (red node). The darker green color represents reference sequences predicted experimentally. The size of the dots at the center of the nodes is proportional to bootstrap values ranging between 50 and 100. **C** Superposition of an eVip structure, predicted by X-ray diffraction (green), and the structural models of an asVip, archaeal viperin (arVip), and bacterial viperin (from left to right). The yellow color in the models emphasizes the high conservation of the viperin catalytic site across the tree of life. The information regarding bacteria, archaea, asgard archaea and eukaryotes in panels (**A**–**C**) are represented by the pink, blue, purple and green color respectively. **D** Anti-T7 phage activity of asVip in *E. coli*. Nine asVip (asVip 26,11,20,25,16,12,17,23,8) exhibited anti-viral activity as indicated by the *p*-values (*$p < 0.05$; **$p < 0.01$). **E** Anti-T7 phage activity of asVip after codon optimization for their expression in *E. coli*. One asVip from a Hodarchaeales organism provided protection against viral infection (asVip 19). The center line of each box plot denotes the median; the box contains the 25th to 75th percentiles. Black whiskers mark the 5th and 95th percentiles. pVip34 is a prokaryotic viperin selected as a positive control from Bernheim et al.[13]. Each experimental condition includes, on average, 53 plaques pooled from three biological replicates. A two-tailed t-test was used to calculate statistical significance in figures (**E**, **D**).

topological arrangement lends strong support to the hypothesis that an early diversification of asAgo gave rise to the wide array of argonaute proteins observed across the tree of life, with occasional instances of horizontal gene transfer (HGT).

The predicted structure of argonautes revealed a clear conservation between asAgo and eAgo (Fig. 3B; red nodes). Detailed examination of five asAgo most closely related to eAgo revealed that three of the proteins initially identified as long types are actually short Agos due to the absence of the PAZ domain (asAgo 2, 4, and 5). The remaining two (asAgo 6 and 7) possess only an isolated PIWI domain, a feature previously reported in approximately 1% of short Agos[28]. Intriguingly, four of these five proteins (asAgo 2, 4, 6, 7) contain a region at the N-terminal portion that is homologous to the eukaryotic translation initiation factor 2 (eIF2). This protein is a member of the PIWI superfamily and is seldom found in prokaryotes. eIF2 plays a critical role in delivering methionyl-tRNA to the 40 S ribosomal subunit and subsequently binding to the 5′ end of capped eukaryotic mRNAs in conjunction with other eIFs[29,30]. Bioinformatics analysis shows a higher occurrence of eIF2 sequences in asAgo compared to other pAgos (Supplementary Data 4). This suggests the possibility of a ribosomal

https://doi.org/10.1038/s41467-024-50195-2

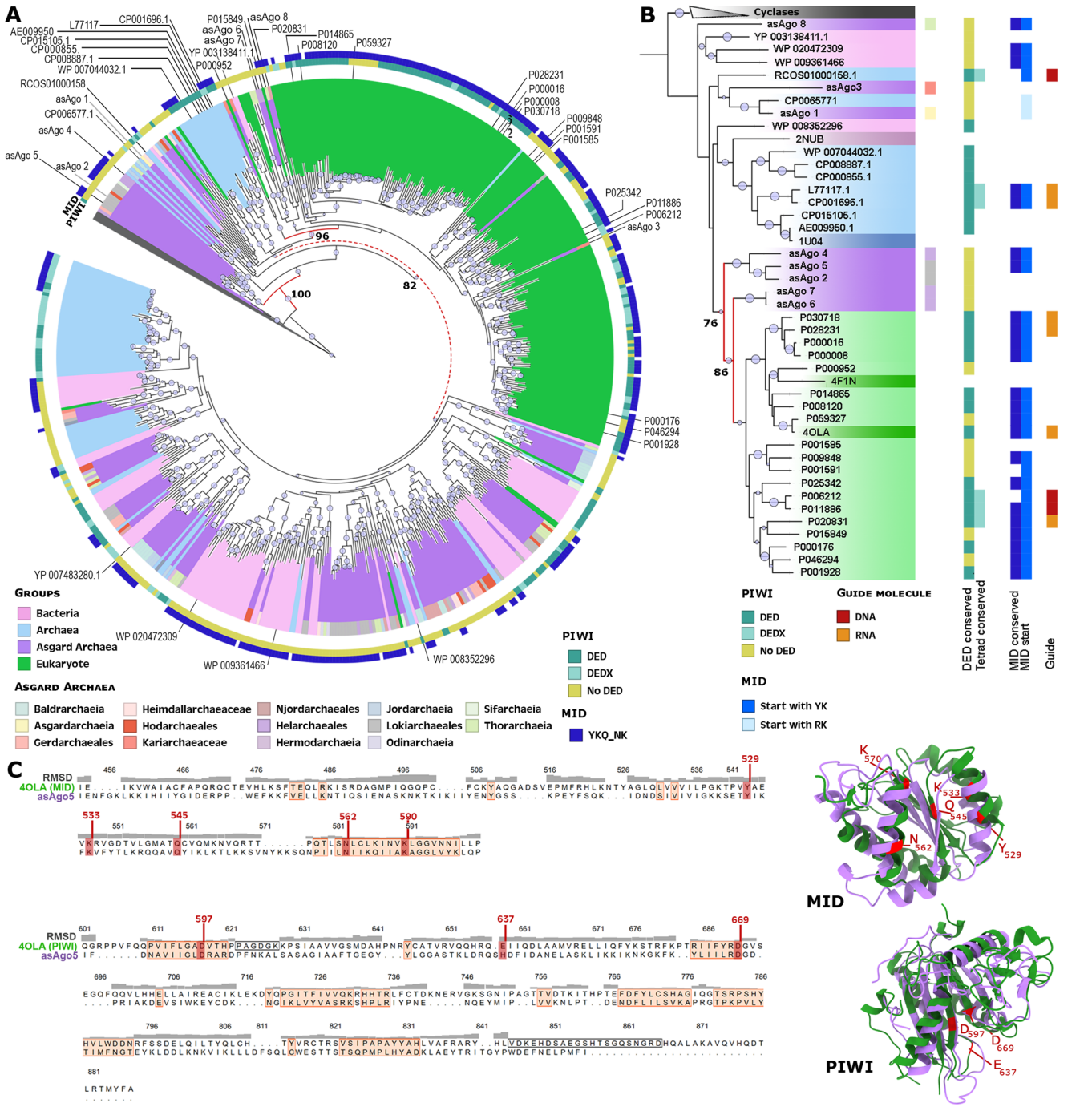

Fig. 3 | Evolutionary history of Asgard argonaute proteins. A Phylogeny of long type argonaute proteins from archaea, bacteria, and eukaryotes with cyclases as outgroup (grey). **B** Structure-based homology of argonautes. **C** Structural alignment of asAgo5 and 4OLA (eAgo) MID and PIWI domains (left), and the graphic model of the corresponding alignments (right). Salmon regions on the alignment highlight strong conservation (low RMDS values). Red amino acids in the structural alignment, and their respective models represent the 4OLA conserved functional residues in MID and PIWI. The information regarding bacteria, archaea, asgard archaea and eukaryotes are represented by the pink, blue, purple and green color respectively. The size of the dots on the nodes is proportional to bootstrap values ranging between 70 and 100.

architecture and eukaryotic-like translation processing in these archaea. This is a hypothesis that still needs experimental confirmation.

Only 20% of the asAgo proteins contain DED catalytic triads, previously thought to be essential for PIWI domains to execute their nuclease function[31]. Of these asAgos (32 in total), 28% have the full tetrad DEDX of amino acids (X = D, H, N or K) (Fig. 3; green and yellow squares) recently shown to be the set of residues necessary for maintaining PIWI activity[32]. These numbers are in accordance with the overall percentage of long pAgo predicted to be active. However, these findings challenge the initial assumption that the ancestral pAgo was an active nuclease[33].

Structural alignment between asAgo and eAgo reveals a high degree of conservation in the PIWI and MID protein domains, even among those asAgos lacking a conserved PIWI tetrad (Fig. 3C). This suggests that ancient asAgo enzymes contained all the components necessary for PIWI nuclease function but lacked the catalytic site. This pre-existing conserved protein organization likely facilitated the

https://doi.org/10.1038/s41467-024-50195-2

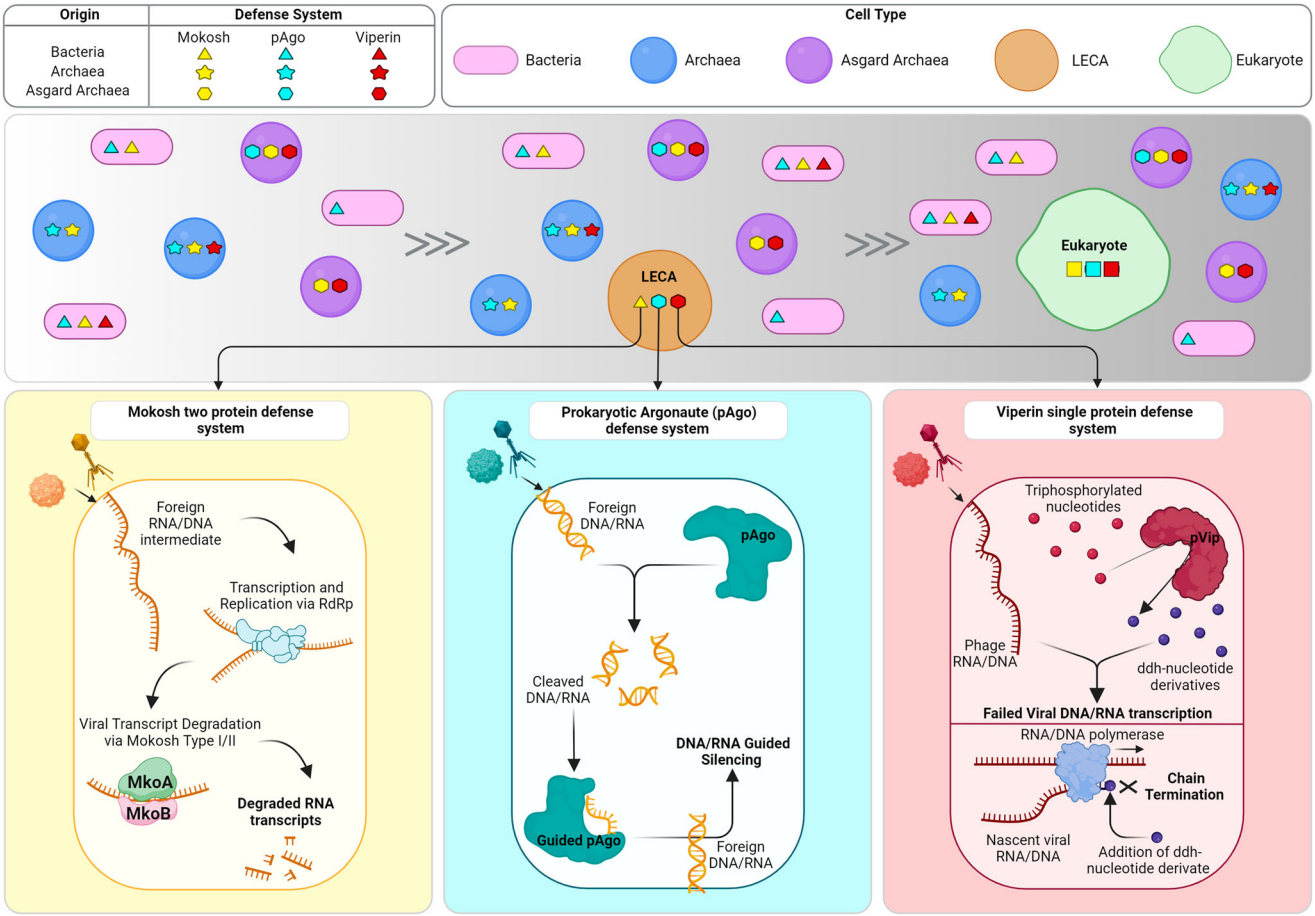

Fig. 4 | Model illustrating the contributions of archaea to the origins of eukaryotic defense mechanisms. Based on our reconstructions of structures and phylogeny in the study, we propose that early forms of both the viperin and argonautes defense systems were inherited from Asgardarchaeota via the LECA (last eukaryotic common ancestor) into modern eukaryotes. However, other systems present in Asgard-like Mokosh appear to have originated from bacteria, highlighting that both prokaryotes have an important role on the origin of defense mechanisms in Eukaryotes. The known mechanisms of mokosh, argonaute, and viperin defense systems are illustrated at the bottom. Created using BioRender.com, released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

perpetuation of this basic structure over time, paving the way for its evolution toward an active nuclease enzyme. Support for this hypothesis comes from the higher number of asAgos with a conserved MID domain binding pocket (YKQ-NK). Experimental evidence has shown that the presence of tyrosine as the first amino acid in the pocket is crucial for a stable binding of the guide nucleotide to argonautes. This is due to threonine's aromatic ring, which functions as a cap to the first base on the 5′ end of the guide molecule[28].

Recently, experimental evidence revealed the function and mechanism of action of an asAgo from a Lokiarchaeales genome. Bastiaanssen et al. characterized the RNA-guided/RNA silencing capabilities of asAgo, suggesting it as a possible origin of this eukaryotic feature before eukaryogenesis[34]. This serves as a notable example of early functionalization of asAgo, where a new function emerges through the adaptation of pre-existing conserved domains.

The single-protein DS known as NLR, named after mammalian Nod-like receptors, serves as an intracellular sensor through its NACHT module, which is responsible for recognizing specific molecular signatures. Recent research suggests that eukaryotes acquired NLR via horizontal gene transfer from bacterial sequences[35]. Our phylogenetic analysis revealed a correlation between the NLR proteins in archaea and eukaryotes (Supplementary Fig. 2A), suggesting they share a common ancestor. Moreover, certain groups of eukaryotic NACHT proteins appear to have a shared ancestry with Asgardarchaeota proteins (Supplementary Fig. 2A; red node). However, more comprehensive analyses and additional NLR-related proteins are required to enhance the resolution of this analysis.

Another DS of interest is Mokosh, which degrades foreign viral transcripts to protect cells from infections[7]. Mokosh system proteins (MkoA and MkoB) have domain homology to the anti-transposon piRNA (PIWI-interacting RNA) pathway in eukaryotes[4]. Our phylogenetic analysis of MkoA suggests a bacterial origin of this protein. However, MkoA homologs associated with defense functions in eukaryotes are distributed in other regions of the phylogenetic tree. Thus, further analyses are needed to elucidate the relationship between functional eukaryotic immune proteins and MkoA homologs (Supplementary Fig. 2B). Like much of the history of cellular complexity in eukaryotes, these findings suggest that eukaryotic DS were derived from both bacteria and archaea (Fig. 4).

This study sheds light on the roles of archaea in the origin of innate immune mechanisms in eukaryotes. Asgardarchaeota emerges as a key player in the origins of viperins and argonautes, with ancestral forms of these proteins tracing back to the last eukaryotic common ancestor (LECA) and prior (Fig. 4). We found remarkable conservation observed in the sequences, structure, and function of these DS proteins compared to their eukaryotic counterparts. Models of eukaryogenesis involve the interaction between a bacterium and an archaeon; our findings reveal that early eukaryotes inherited robust defense mechanisms against viral infections from both ancestral partners (Fig. 4). The compatibility of some of these proteins (asVips

and asAgos) with those in eukaryotes suggests these systems may provide biomedical or biotechnology applications in eukaryotes.

Methods

Known defense systems identification

A database of Archaeal genomes available from NCBI was built containing 646 genomes from the DPANN group, 668 genomes from TACK superphylum, 1408 genomes belonging to Euryarchaeota phylum, and 869 Asgard archaea genomes[36]. This database only contains complete or high-quality genomes. To guarantee the quality of the genomes, a CheckM v1.2.1 screening was performed and only genomes presenting a completeness higher than 45%, and contamination lower than 10% were included. Another database containing 716 bacterial genomes from 71 different phyla was built following the same procedure. The distribution and diversity of known defense systems in archaeal genomes was checked running DefenseFinder v1.0.9[6] with models v1.2.2. and default parameters on the custom databases.

Selecting homologues of archaeal defense system proteins in Eukaryotic genomes

A database of defense system proteins homologues found in Eukaryotes created by Cury et al.[4] was used as a template to search for homologues of archaeal defense system associated proteins. A DIAMOND v2.0.13.151[37] database was created with Cury's dataset and 4 separated queries were prepared, one with proteins from each archaeal group (DPANN, Eury, TACK and Asgard) (Supplementary Data 1). A pairwise alignment was performed with DIAMOND v2.0.13.151 using BLASTP with a cutoff of >30% identity and $<E^{-05}$ e-value. In addition, DefenseFinder was run on genomes from the EukProtV3 database[38] to search for additional homologues of prokaryotic defense systems in eukaryotes following the same parameters described in the section above. In Fig. 1A bubble plot, the less frequent DS in Eukaryotic genomes (31 in total: RTnitrilase-Tm, Old Tim, Hydrolase-3Tm, HelicaseDUF2290, gop beta cll, DUF4238, DprA-PPRT, 3HP, 2TM 1TM TIR, RosmerTA, PD-T7-5, PD-T7-1, PD-T4-3, PD-T4-2, PD-T4-1, Gao Tmm, Gao TerY, Gao RL, Gao Qat, Gao Ppl, Gao Mza, Gao let, Gao Hhe, Gao Her, RnlAB, Old exonuclease, Nhi, PD-T4-6, PD-T4-5, AbiT, AbiQ) were excluded from the visual representation, but can have their distribution check in each dataset on Supplementary Data 1.

Phylogenetic analyses

The viperins phylogenetic analysis is shown in Fig. 3. It contains viperin protein sequences identified by DefenseFinder in the archaeal database together with the viperins previously characterized and used on the phylogenetic analyses on Bernheim et al.[13] work. After removing redundant sequences with seqkit v.2.3.036, a total of 337 unique sequences were processed as described in the end of this section to generate a dendrogram (model selected: LG + R10).

To analyze the homology relationship between argonaute sequences, we combined the Long pAgo and eAgo sequences used by Swarts et al.[33] with COG1431_pAgo and pAgo_LongA sequences identified in our archaea database and the EukProtV3 database[38] with DefenseFinder. Identical sequences were removed using seqkit v.2.3.0[39], resulting in a total of 543 unique argonaute proteins. The sequences were aligned using MAFFT v7.457[40]. Regions in the alignment containing > 90% gaps were removed using ClipKIT 1.3.0[41]. The tree was computed with IQTree v2.0.3 using 1000 bootstrapping replicates, and the best model generated (LG + F + R10) was selected according to the Bayesian information criterion (BIC)[42]. The same alignment, gap removal, and dendrogram generation procedures were used for all the protein sequence phylogenetic analyses in this section. All information about the protein sequences used in these analyses can be found in Supplementary Data 4. Alignments used to generate the phylogenetic analyses can be found at https://doi.org/10.6084/m9.figshare.25838197 (Supplementary Data 5).

Structural homology analyses

To analyze the homology between proteins used as defense mechanisms in prokaryotes and eukaryotes, protein sequences associated with defense systems were submitted to structural modeling in ColabFold v1.5.2[43]. Multi-sequence alignments were performed using the mmseq2 mode, and AlphaFold2_ptm model. Three recycling steps were used to optimize the computational power and running time.

The predicted protein models were combined with reference structures of the respective proteins acquired from RCSB Protein Data Bank (RCSB). A multi-structural alignment (MSTA) of these structures was performed using the default parameters of mTM-align v.20220104 tool[44] to build a dendrogram using IQTree v2.0.3 (models: LG + I + G4 for viperins; VT + R4 for argonautes).

The pairwise RMSD matrix obtained from mTM-aligns (Supplementary Data 3) was used to select the proteins suitable for the 3D reconstruction of their alignment using the Needleman-Wunsch algorithm and BLOSUM-62 matrix on ChimeraX software v.1.7.1[45].

When necessary, protein annotation was done using Interproscan v.5.31-70.0[46] with the default parameters (Supplementary Data 4). Alignments used to generate the structural analyses can be found at https://doi.org/10.6084/m9.figshare.25838197 (Supplementary Data 5).

Bacterial growth and phage propagation

E. coli strains (MG1655, Keio ΔiscR27, DH5α) were grown in LB or LB agar at 37 °C. Whenever applicable, media were supplemented with chloramphenicol (25 μg mL^{-1}) and/or kanamycin (50 μg mL^{-1}) to maintain the plasmids. T7 phage was propagated on *E. coli* BL21s using the plate lysate method. Lysate titre was determined using the small drop plaque assay method as previously described[47].

Plasmid construction

All primers were purchased from IDT (Supplementary Data 2). asVip genes were synthesized by Twist Biosciences, and codon optimized where indicated. Genes were sub-cloned into an inducible expression vector using Golden Gate assembly. All plasmids were propagated in *E. coli* DH5α and then purified and transformed into the Keio ΔiscR27 strain to upregulate iron-sulfur cluster production, as described previously[13,48].

Plaque assays

Plaque assays were performed according to standard protocols[47]. Overnight cultures were diluted 1:100 in LB supplemented with 1.25 mM $MgCl_2$, 1.25 mM $CaCl_2$ and Isopropyl β-D-1-thiogalactopyranoside (IPTG; final concentration of 0.5 mM) for induction of asVip expression and grown to an OD_{600} ~0.3. Bacteria from these outgrowth cultures were combined with serial dilutions of phage lysate and incubated at 37 °C for 15 min before being mixed with 0.5% agar and plated on agar plates. After solidifying, the plates were incubated overnight at 37 °C. Plaques were imaged using an Azure Biosystems 600 imaging system, and their areas were measured using FIJI's 'Analyze Particles' plugin[49]. Three biological replicates were pooled for analysis. A two-tailed t-test was used to calculate statistical significance.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The final assembled and annotated asgard genomes used in this study are available in NCBI under BioProjects PRJNA743900, PRJNA692327, PRJNA1112871. All the other defense system proteins used in this study can have their accession number recovered from the respective supplementary information files. The raw data used to generate the figures

presented in this work, and supplementary file can be found here: https://doi.org/10.6084/m9.figshare.25838197.

References

1. Mayo-Muñoz, D., Pinilla-Redondo, R., Birkholz, N. & Fineran, P. C. A host of armor: Prokaryotic immune strategies against mobile genetic elements. *Cell Rep.* **42**, 112672 (2023).
2. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *elife* **4**, e08490 (2015).
3. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
4. Cury, J. et al. Conservation of antiviral systems across domains of life reveals novel immune mechanisms in humans. Preprint at *bioRxiv* https://doi.org/10.1101/2022.12.12.520048 (2022).
5. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
6. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
7. Millman, A. et al. An expanded arsenal of immune systems that protect bacteria from phages. *Cell host-microbe* **30**, 1556–1569 (2022).
8. Eme, L. et al. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 1–8 (2023).
9. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
10. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
11. Chin, K. C. & Cresswell, P. Viperin (cig5), an IFN-inducible antiviral protein directly induced by human cytomegalovirus. *Proc. Natl Acad. Sci.* **98**, 15125–15130 (2001).
12. Fenwick, M. K., Li, Y., Cresswell, P., Modis, Y. & Ealick, S. E. Structural studies of viperin, an antiviral radical SAM enzyme. *Proc. Natl Acad. Sci.* **114**, 6806–6811 (2017).
13. Bernheim, A. et al. Prokaryotic viperins produce diverse antiviral molecules. *Nature* **589**, 120–124 (2021).
14. Meeks, J. C. Physiological adaptations in nitrogen-fixing Nostoc–plant symbiotic associations. In *Prokaryotic Symbionts in Plants* (pp. 181–205). Berlin, Heidelberg: Springer Berlin Heidelberg (2007).
15. Li, F. W. et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. plants* **4**, 460–472 (2018).
16. Lachowicz, J. C., Gizzi, A. S., Almo, S. C. & Grove, T. L. Structural insight into the substrate scope of viperin and viperin-like enzymes from three domains of life. *Biochemistry* **60**, 2116–2129 (2021).
17. Shomar, H. et al. Viperin immunity evolved across the tree of life through serial innovations on a conserved scaffold. *Nature Ecology & Evolution*, 1–13 (2024).
18. Rivera-Serrano, E. E. et al. Viperin reveals its true function. *Annu. Rev. Virol.* **7**, 421–446 (2020).
19. Penev, P. I. et al. Supersized ribosomal RNA expansion segments in Asgard archaea. *Genome Biol. Evol.* **12**, 1694–1710 (2020).
20. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
21. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293–296 (2000).
22. Cerutti, L., Mian, N. & Bateman, A. Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem. Sci.* **25**, 481–482 (2000).
23. Shabalina, S. A. & Koonin, E. V. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587 (2008).
24. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 1–15 (2009).
25. Willkomm, S., Makarova, K. S. & Grohmann, D. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. *FEMS Microbiol. Rev.* **42**, 376–387 (2018).
26. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
27. Sheu-Gruttadauria, J. & MacRae, I. J. Structural foundations of RNA silencing by argonaute. *J. Mol. Biol.* **429**, 2619–2639 (2017).
28. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The expanded universe of prokaryotic Argonaute proteins. *MBio* **9**, 10–1128 (2018).
29. Sachs, A. B., Sarnow, P. & Hentze, M. W. Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* **89**, 831–838 (1997).
30. Erickson, F. L., Nika, J., Rippel, S. & Hannig, E. M. Minimum requirements for the function of eukaryotic translation initiation factor 2. *Genetics* **158**, 123–132 (2001).
31. Wang, S. et al. Electron transport chains in organohalide-respiring bacteria and bioremediation implications. *Biotechnol. Adv.* **36**, 1194–1206 (2018).
32. Wang, X. et al. Structural insights into mechanisms of Argonaute protein-associated NADase activation in bacterial immunity. *Cell Res.* **33**, 1–13 (2023).
33. Swarts, D. C. et al. The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014).
34. Bastiaanssen, C. et al. RNA-guided RNA silencing by an Asgard archaeal Argonaute. *Nat Commun* **15**, 5499 (2024).
35. Kibby, E. M. et al. Bacterial NLR-related proteins protect against phage. *Cell* **186**, 2410–2424 (2023).
36. Appler, K. E. et al. Oxygen metabolism in descendants of the archaeal-eukaryotic ancestor. Preprint at bioRxiv https://doi.org/10.1101/2024.07.04.601786v1 (2024).
37. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. methods* **18**, 366–368 (2021).
38. Richter, D. J. et al. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community J.* **2**, e56 (2022).
39. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS one* **11**, e0163962 (2016).
40. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
41. Steenwyk, J. L. et al. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, e3001007 (2020).
42. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
43. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. methods* **19**, 679–682 (2022).
44. Dong, R., Peng, Z., Zhang, Y. & Yang, J. mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* **34**, 1719–1725 (2018).
45. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
46. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
47. Clokie, M. R. & Kropinski, A., 2009. Methods and protocols, volume 1: Isolation, Characterization, and Interactions. *Methods Mol. Biol.* Humana press, pp.69–81.

48. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006–0008 (2006).
49. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. methods* **9**, 676–682 (2012).

Acknowledgements

This work was supported by the Moore-Simons Project on the Origin of the Eukaryotic Cell, Simons and Moore Foundations 73592LPI to B.J.B. (https://doi.org/10.46714/735925LPI) and Welch Foundation (F-1808) to I.J.F.

Author contributions

P.L., B.J.B. and V.D.A designed the study. B.J.B. and I.J.F. supervised and provided infrastructure to execute this study. K.E.A. and V.D.A. performed metagenomic data assemblage and binning, providing MAGs for the study. P.L., E.A.P., K.C., and V.D.A built and curated the datasets used in this study. M.E.L., D.S., and I.J.F. were responsible for the design and execution of viperins experimental analyses. P.L., E.A.P. and K.C. performed phylogenetic analyses. P.L. executed the protein structural homology analyses. P.L., M.E.L., and B.J.B. wrote the manuscript. All authors contributed to the final version of the manuscript and approved it before submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-50195-2.

Correspondence and requests for materials should be addressed to Pedro Leão or Brett J. Baker.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.