Sexual reproduction consists of genome reduction by meiosis and subsequent gamete fusion. The presence of genes homologous to eukaryotic meiotic genes in archaea and bacteria suggests that DNA repair mechanisms evolved towards meiotic recombination. However, fusogenic proteins resembling those found in gamete fusion in eukaryotes have so far not been found in prokaryotes. Here, we identify archaeal proteins that are homologs of fusexins, a superfamily of fusogens that mediate eukaryotic gamete and somatic cell fusion, as well as virus entry. The crystal structure of a trimeric archaeal fusexin (Fusexin1 or Fsx1) reveals an archetypical fusexin architecture with unique features such as a six-helix bundle and an additional globular domain. Ectopically expressed Fusexin1 can fuse mammalian cells, and this process involves the additional globular domain and a conserved fusion loop. Furthermore, archaeal fusexin genes are found within integrated mobile elements, suggesting potential roles in cell-cell fusion and gene exchange in archaea, as well as different scenarios for the evolutionary history of fusexins.
How early eukaryotes developed the capacity for gamete fusion is a central question entangled with the origin of the eukaryotic cell itself. The widespread presence of a conserved set of meiosis, gamete, and nuclear fusion proteins among extant eukaryotes indicates that meiotic sex emerged once, predating the last eukaryotic common ancestor (LECA)\textsuperscript{1–2}. Two essential molecular events are required for meiotic sex: DNA recombination and plasma membrane fusion. Fused by the crystallographic structures of the ectodomain of EFF-1 demonstrated (sexual fusexins; hereafter referred to as HAP2) that mediate gamete fusion between heterologous cells, and substitute a viral fusogen because EFF-1 and AFF-1 use a bilateral mechanism while class II fusogens use a unilateral mechanism\textsuperscript{13}. However, their mechanisms of membrane fusion are different to mediate plasma membrane fusion only when these fusogenic proteins are expressed in both fusing membranes\textsuperscript{13–15}. The crystal structure of the ectodomain of EFF-1 demonstrated structural similarity with class II viral fusogens\textsuperscript{12}. EFF-1 and AFF-1 can fuse cells in C. elegans, promote fusion between heterologous cells, and substitute a viral fusogen to mediate plasma membrane fusion only when these fusogenic proteins are expressed in both fusing membranes\textsuperscript{13–15}. The crystal structure of the ectodomain of EFF-1 demonstrated structural similarity with class II viral fusogens such as the glycoproteins on the surface of Zika, dengue, and rubella viruses\textsuperscript{16}. Thus, these eukaryotic and viral fusogens have remarkably similar functions and structures despite undetectable sequence similarity. However, their mechanisms of membrane fusion are different because EFF-1 and AFF-1 use a bilateral mechanism while class II viral fusogens use a unilateral mechanism\textsuperscript{13–15}.

More recently, genetic screens uncovered a protein involved in gamete fusion, HAP2/GCS1, which is conserved in Arabidopsis, Chlamydomonas, Plasmodium, Tetrahymena, and Dictyostelium\textsuperscript{17–21}. Later, structural bioinformatics, crystallographic structure elucidation and functional assays demonstrated that HAP2/GCS1 is a bona fide fusogen homologous to EFF-1, AFF-1, and class II viral fusogens\textsuperscript{22–24}. The crystal structures of HAP2/ GCS1 from Chlamydomonas and Arabidopsis showed remarkable structural conservation without sequence similarity\textsuperscript{23,25,26}. This superfamily of fusion proteins was named fusexins: fusion proteins essential for sexual reproduction and exoplasmic merger of plasma membranes\textsuperscript{22–24}. Thus, the fusexin superfAMILY encompasses class II viral fusogens (viral fuseXins) that fuse the envelope of some animal viruses with the membranes of host cells during infection\textsuperscript{34}; EFF-1 and AFF-1 (somatic fuseXins) that promote cell fusion during syngamy organ development\textsuperscript{10,13,15,16}; and HAP2/GCS1 (sexual fuseXins; hereafter referred to as HAP2) that mediate gamete fusion\textsuperscript{17–19}.

Although it is assumed that sexual fuseXins were already present in the LECA\textsuperscript{1,27}, their shared ancestry with viral fuseXins posed a “the virus or the egg” evolutionary dilemma\textsuperscript{22,24,28}. In one scenario, fuseXins are proper eukaryal innovations that were captured by some viruses and used for host invasion. Alternatively, a viral fusexin gene was transferred to an early eukaryotic cell and then repurposed for gamete fusion. Solving this evolutionary conundrum is not a trivial task because sequence-based phylogenetics cannot be applied to the whole fusexin superfAMILY due to lack of sequence conservation.

Here we identify a family of fuseXins in genomes of Archaea and prokaryotic fractions of metagenomes from very diverse environments. We provide crystallographic and functional evidence suggesting that these proteins are cellular fusexins. Genomic analyses show that archaeal fuseXins are carried by integrated mobile genetic elements. Evolutionary analyses of the whole fusexin superfAMILY reveal alternative working models regarding the relationships between viral, eukaryotic and archaeal fuseXins and the emergence of meiotic sex during eukaryogenesis.

**Results**

**Fusexin genes in Archaea.** To search for fuseXins we used the crystallographic structures of C. reinhardtii, A. thaliana, and T. cruzi HAP2 (Cr/At/TcHAP2)\textsuperscript{23,25,26} to build dedicated Hidden Markov Models (HMMs) for scanning the Uniclust30 database. We detected 24 high-confidence candidates in prokaryotes: 8 belong to isolated and cultivated archaea, and the remaining 16 to metagenome-assembled genomes (MAGs, Supplementary Table 1). We then built HMMs of the candidate ectodomains and compared them to HMMs of sexual, somatic, and viral fuseXins. Figure 1a shows that the prokaryotic candidates have detectable sequence similarities with HAP2, with E-values below 0.001 and HHblits-derived probabilities higher than 0.95 (Supplementary Fig. 1a). We named these proteins Fusexin1 (Fsx1). Fsx1 genes found in cultivated and isolated prokaryotes are restricted to the Haloarchaea class (also called Halobacteria, Euryarchaeota superphylum) whereas MAGs containing Fsx1s include all major Archaea superphyla (Supplementary Table 1). Next, we used this Fsx1 sequence set to search the Metaclust database, which comprises 1.59 billion clustered proteins from over 2200 metagenomic/metatranscriptomic datasets. Performing a scan pipeline using PHMMER, PSI-BLAST, HMM–HMM comparisons and topology filtering we found 96 high-confidence fsx1 genes. The identified fsx1s come from different environments (with predominance of saline samples) and a wide temperature range (−35 to 80 °C, see Supplementary Data 1).

Fsx1 is a structural homolog of HAP2/GCS1. To experimentally investigate the presence of fusexin-like proteins in Archaea, a selection of the candidate genes was expressed in mammalian cells (Supplementary Fig. 1b, c). High-level expression was observed for a metagenomic Fsx1 sequence from a hypersaline environment, predicted to encode a ~55 kDa ectodomain region (Fsx1\textsubscript{E}) followed by three transmembrane domains (TM) (Supplementary Data 1). Fsx1\textsubscript{E} is a monomer in solution but crystallized as a homotrimer in the presence of 2.5 M NaCl, 0.2 M CaCl\textsubscript{2} (Supplementary Fig. 2). These conditions precluded experimental phasing, and attempts to phase the data by molecular replacement (MR) with different kinds of homology models also failed, due to insufficient sequence identity to known fusexin structures. However, we could determine the structure of Fsx1\textsubscript{E} at 2.3 Å resolution by running MR with a combination of fragments from ab initio predictions generated by AlphaFold\textsuperscript{29} (Fig. 1b, Supplementary Figs. 3, 4 and Supplementary Table 2).

The Fsx1\textsubscript{E} homotrimer has overall dimensions of 119 × 77 × 76 Å (Fig. 1b). Each protomer consists of four domains (Fig. 2a, b), the first three of which match the approximate dimensions and relative arrangement of domains I–III of known fuseXins in their post-fusion conformation\textsuperscript{30}; accordingly, fold...
Fig. 1 Fsx1 is a member of the fusexin protein superfamily. a HMM vs. HMM homology probabilities of a subset of eukaryotic, viral, and archaeal fusexin ectodomains. With the exception of Fsx1, which derives from a metagenomic sequence, archaeal fusexins (red), viral fusexins (blue), EFF-1 (green) and HAP2s (black) are indicated by RefSeq/UniProt identifiers. b Crystal structure of the trimeric ectodomain of Fsx1. The three-fold non-crystallographic axis is indicated. Subunit A is shown as a cartoon colored by domain, with disulfides and the fusion loop (FL) colored magenta and orange, respectively; subunits B, C are in mixed cartoon/surface representation.

and interface similarity searches identify HAP2 as the closest structural homolog of Fsx1B, followed by viral fusexins and C. elegans EFF-1 (Fig. 2c, d). Fsx1 domains I and III are relatively sequence-conserved among archaeal homologs (Supplementary Figs. 5a and 6) and closely resemble the corresponding domains of HAP2 (RMSD 2.1 Å over 218 Ca), including the invariant disulfide bond between domain III strands βC and βF23 (C389-C432; Fig. 2, Supplementary Fig. 4c). On the other hand, Fsx1 domain II shares the same topology as that of HAP2 but differs significantly in its secondary structure elements and their relative orientation, as well as disulfide bonds (Fig. 2c). In particular, Fsx1 domain II is characterized by a four-helix hairpin, whose N-terminal half interacts with the same region of the other two subunits to generate a six-helix bundle around the molecule’s three-fold axis (Figs. 1b and 3a–c and Supplementary Figs. 4a and 5b, c).

Notably, unlike previously characterized viral and eukaryotic fusexins, Fsx1 also contains a fourth globular domain conserved among archaeal homologs (Figs. 1b, 2 and Supplementary Figs. 4d and 6). Its antiparallel β-sandwich, which includes the two C-terminal disulfides of Fsx1, resembles the carbohydrate-binding fold of dust mite allergen Der p 23 and related chitin-binding proteins31 (Fig. 3d); accordingly, it is also structurally similar to a high-confidence AlphaFold2 model of the C-terminal domain of acidic mammalian chitinase32. In addition to being coaxially stacked with domain III as a result of a loop/loop interaction stabilized by the C457-C477 disulfide, domain IV contributes to the quaternary structure of the protein by interacting with domain II of the adjacent subunit to which domain III also binds (Figs. 1b and 3c).

The Fsx1L monomer has a net charge of −67, and another feature stabilizing its homotrimeric assembly is a set of Ca2+ and Na+ ions that interacts with negatively charged residues at the interface between subunits (Figs. 3a–c and Supplementary Figs. 4a and 5b). Additional metal ions bind to sites located within individual subunits; in particular, a Ca2+ ion shapes the conformation of the domain II cd loop (S143-V148) so that its uncharged surface protrudes from the rest of the molecule (Fig. 3e, f and Supplementary Fig. 5d). Strikingly, the position of this element matches that of the fusion loops (FLs) of other fusexins, including the Ca2+-binding fusion surface of rubella virus E1 protein33,34 (Fig. 3e). Moreover, as previously observed in the case of CrHAP226, the loops of each trimer interact with those of another trimer within the Fsx1 crystal lattice.

In summary, despite significant differences in the fold of domain II, the unprecedented presence of a domain IV and extreme electrostatic properties, the overall structural similarity between Fsx1 and viral or eukaryotic fusexins suggests that this prokaryotic molecule also functions to fuse membranes.

Fsx1 can fuse eukaryotic cells. To test the fusogenic activities of the candidate archaeal fusexins we studied their fusion activity upon transfection in eukaryotic cells15,16,22. Cells with either red or green nuclei are mixed with each other and fusion is measured by the formation of hybrid cells with both red and green nuclei revealing merger of their cytoplasm. For this, we co-cultured two batches of Baby Hamster Kidney (BHK) cells independently transfected with Fsx1 and co-expressing either nuclear H2B-RFP or H2B-GFP22. We then performed immunofluorescence against a V5 tag fused to the cytoplasmic tail of Fsx1 (Fig. 4a, b, and Supplementary Fig. 7a, b). We observed a five-fold increase in the mixing of the nuclear H2B-GFP and H2B-RFP compared to vector control, showing that Fsx1 is a bona fide fusogen, as efficient as the fusexin AtHAP2 (Fig. 4c). To determine whether Fsx1 expression is required in both fusing cells or, alternatively, it suffices in one of the fusing partners, we mixed BHK-Fsx1 coexpressing cytoplasmic GFP with BHK cells expressing only nuclear RFP. We found increased multinucleation of GFP coexpressing cytoplasmic GFP with BHK cells expressing only nuclear GFP. We found increased multinucleation of GFP+ cells (revealing cell–cell fusion) but very low mixing with RFP+ cells not expressing Fsx1. In contrast, the vesicular stomatitis virus G-glycoprotein (VSVG) fusogen-induced efficient unilateral fusion15 (Fig. 4d–f). While VSVG requires acidic pH for maximum fusion activity Fsx1-mediated multinucleation was not stimulated by low pH (Supplementary Fig. 1e). Thus, Fsx1 acts in a bilateral way, similarly to EFF-1 and AFF-1 fusexins14,15,35. We then performed live imaging using spinning disk confocal microscopy and observed bilateral cell-cell fusion of BHK-Fsx1 cells (Fig. 4g, h).
Fig. 2 Domain architecture of Fsx1 and topological comparison with HAP2. a Schematic diagram of the domains of Fsx1. SP, signal peptide; TM, predicted transmembrane helices. Note that Fsx1 is predicted to contain three C-terminal TMs, with a cytoplasmic loop between TM1 and TM2 that lacks Cys residues; on the other hand, HAP2 homologs are characterized by having a single TM, followed by a cytoplasmic tail that often contains Cys implicated in fusion64. b Crystal structure of the ectodomain of Fsx1 and predicted topology of the full-length protein relative to the plasma membrane. Domains I–IV are shown in red, yellow, blue, and green, respectively; disulfide bonds are indicated and colored magenta. c Topology diagrams of the ectodomains of Fsx1, C. reinhardtii HAP2 (PDB 6E18 [https://www.rcsb.org/structure/6E18])65, A. thaliana HAP2 (PDB 5OW3 [https://www.rcsb.org/structure/5OW3])25, and T. cruzi HAP2 (PDB 5OW4 [https://www.rcsb.org/structure/5OW4])25. Domains and disulfide bonds are colored as in panel b. Note how, although domain II of Fsx1 has the same topology as the corresponding domain of HAP2, it contains only one of its conserved disulfide bonds (C1125–C2155, corresponding to C147-C210/disulfide bond 3 of CrHAP223). d Side-by-side comparison of Fsx1 E and known class II fusogens. Elements are colored as in panel b; the stem region and the linker between domains I and III are shown in pink and cyan, respectively.
Structure–function analysis of Fsx1. To compare archaeal Fsx1 activity with fusexins from eukaryotes and viruses, we introduced mutations into three different structural domains of Fsx1 and tested surface expression and fusogenic activity in mammalian cells.

First, to test whether the putative FL of Fsx1 (143-SVTSPV-148) is involved in fusion, we replaced it with a linker of 4G between Y142A and Y149A (Figs. 3e, 5a, and Supplementary Fig. 5d). This replacement does not affect surface expression yet reduces cell–cell fusion to levels similar to those of the negative control (Fig. 5b–f).

Second, we asked whether domain IV, which is only present in archaeal fusexins, has a function in the fusion process. For this, we replaced the entire domain with the stem region of EFF-1 (Figs. 3d and 5a; ΔDIV → EFF-1 stem). While this mutant Fsx1 reaches the cell surface, suggesting that it folds normally, it shows a significantly reduced activity compared to wild-type Fsx1 (Fig. 5b–f).

Third, to test whether the three TMs of Fsx1 are required for fusion, we replaced them with the TM and cytoplasmic domains of AtHAP2 (Fig. 5b; ΔTMs → EFF-1 TM) or a glycosylphosphatidylinositol (GPI) anchor signal (Fig. 5a; ΔTMs → GPI). We found that both Fsx1 mutants remained active (Fig. 5b), indicating that the Fsx1 TMs are not essential for fusion. Finally, we also replaced the TM and cytoplasmic domains of AtHAP2 with a signal for GPI and found that the protein also maintained its fusogenic activity (Figs. 4c, and 5a, b). Thus, contrary to some viral fusogens in which the GPI-anchored glycoproteins fail to drive complete fusion36–38, lipid-anchored Fsx1 or eukaryotic HAP2 promote syncytia formation when expressed on the surface of BHK cells.

Fig. 3 Distinct structural features of Fsx1. a Fsx1 surface colored by electrostatic potential from red (−10 kT/e) to blue (+10 kT/e) through white (0 kT/e). Orange arrows indicate the FLs. b Location of the Ca2⁺ and Na⁺ ions (depicted as yellow and purple spheres, respectively; see also Supplementary Figs. 4 and 5) stabilizing the Fsx1 trimer. The molecular surface of a protomer is shaded green and the outline of the other two subunits is colored cyan and wheat. c Details of interactions at the level of the six-helical bundle made by domain II of the Fsx1 subunits (right subpanels) and the domain IV/domain II interface (bottom left subpanel). Selected side chains are colored by the type of inter-chain contacts in which they are involved (gray: hydrophobic interaction; blue bell: hydrogen bonding; yellow: Ca2⁺ coordination; fuchsia: Na⁺ coordination), with dashed lines indicating hydrogen bonds. Note that the helical bundle of Fsx1 is not a leucine-zipper coiled-coil structure, such as those found in class I/III viral fusion proteins or in the SNARE four-helix bundles, and see also Supplementary Fig. 5b. d Superposition of Fsx1 domain IV (green) and Der p 23 (PDB 4ZCE [https://www.rcsb.org/structure/4ZCE] blue) (Dali score 3.6, RMSD 2.2 Å). e Comparison of the Fsx1 region that includes the FL and the corresponding parts of CrHAP2 and Rubella virus E1 protein (PDB 4B3V [https://www.rcsb.org/structure/4B3V]) residues coordinating the Ca2⁺ ion that stabilizes the Fsx1 FL are underlined, and compared to the E1 protein Ca2⁺-binding region in the boxed panels on the far right. f The Fsx1 FL adopts a highly ordered conformation stabilized by a Ca2⁺ ion. Presence and identity of the latter, indicated by a yellow arrow, are supported by two other maps shown in addition to the 2mFo-DFc map (blue mesh, contoured at 1.0 σ); a difference map calculated upon omitting all metal ions from the model (thick green(+) / red(−) mesh, 6.0 σ) and a phased anomalous difference map calculated from a 2.9 Å-resolution dataset collected at 7.1 keV (thick magenta mesh, 3.2 σ).
Fsx1s are ancient fusogens associated with integrated mobile elements. The $fsx1$ genes here identified are present in a wide physicochemical landscape (Fig. 6a). We observed that the branching pattern of Fsx1 sequences from complete genomes is incompatible with their species tree (Fig. 6a, b). This and the sparse pattern of Fsx1 presence in Archaea led us to perform genomic comparisons of related species with and without the $fsx1$ gene. These revealed >50 kbp DNA insertions in the genomes of species with $fsx1$ genes (Fig. 7a). To investigate them, we performed k-mer spectrum analysis on $fsx1$-containing Pure Culture Genomes (PCGs) and found divergent regions containing the $fsx1$ ORF (Fig. 7b). Gene content analyses of $fsx1$-containing regions show that they share a portion of their genes (Supplementary Fig. 9) and display conserved synteny (Fig. 8), suggesting common ancestry. These regions are enriched in ORFs homologous to proteins involved in DNA mobilization and integration (Fig. 8 and Supplementary Table 3). Thus, our results indicate that $fsx1$ genes are contained in integrated mobile elements (IMEs) that can be mobilized by a conjugative-like, cell fusion-dependent mechanism.

Fig. 4 Fsx1 mediates bilateral cell–cell fusion. a–c Cell–cell fusion was measured by content-mixing, indicated by the appearance of multinucleated cells containing green nuclei (H2B-GFP) and magenta nuclei (H2B-RFP). Immunofluorescence against the V5 tag was also performed (gray). a Images of mixed cells. DAPI, blue. Scale bars, 20 μm. See also Supplementary Fig. 7a. b Scheme of experimental design. c Quantification of content-mixing. The mixing indices presented as means ± SEM of four independent experiments. Comparisons by one-way ANOVA followed by Bonferroni’s test. ns = non-significant, ***p < 0.001, ****p < 0.0001. Source data are provided as a Source Data file. d–f Unilateral fusion was evaluated by mixing control cells expressing nuclear H2B-RFP (magenta) with cells expressing GFP with a nuclear export signal (GFPnes, green cytoplasm) only or together with Fsx1, EFF-1 or VSV G. d Images of cells transfected with empty::GFPnes vector, Fsx1::GFPnes, EFF-1::GFPnes or VSV G::GFPnes. Fsx1 and EFF-1 show multinucleated GFPnes positive cells (arrowheads). VSV G multinucleated cells are found with GFP only (arrowheads) or with both markers (arrows). Scale bars, 20 μm. See also Supplementary Fig. 7e. e Scheme of experimental design. f Quantification of content-mixing experiments in which only the GFP population of cells express Fsx1, EFF-1, VSV G, or none of them (vector). Bar chart showing means ± SEM of three independent experiments. Comparisons by one-way ANOVA followed by Dunnett’s test against the vector. ns = non-significant, **p < 0.01, ***p < 0.001, ****p < 0.0001. Source data are provided as a Source Data file. g Spinning disk microscopy time-lapse images indicating the merging of two cells expressing myr-mCherry and Fsx1. Time in hours:minutes. The red channel (mCherry, white) and the DIC are shown. Refer to Supplementary Movie 1. h For the last point a Z-projection showing the myr-mCherry fluorescence (white) and the nuclei Hoechst (blue, Supplementary Movie 2). Scale bar, 20 μm.
To describe Fsx1’s tempo and mode of evolution we first compared archaeal and sexual fusexins, which share enough sequence conservation to apply standard phylogenetic analyses, not possible for somatic and viral fusexins, as these methods are not able to cope with such amount of divergence (Fig. 1a and Supplementary Fig. 1a). We built maximum likelihood (ML) phylogenies for a set of Fsx1 sequences derived from isolated species and metagenomes, and a subset of HAP2s which capture the full phylogenetic diversity present in eukaryotic lineages (Supplementary Fig. 10a). A major finding...
comes from these phylogenies: eukaryal and archaeal fusexins cluster into strongly supported clades suggesting they diverged before LECA.

To place fsx1 in the broader fusexin superfamily context, we performed structural phylogenetic analysis comparing crystal structures from viral, somatic, and eukaryotic gamete fusegosins (Supplementary Fig. 10b). This structure-based tree supports a viral origin of somatic fusexins (EFF-1) and is also compatible with archaeal fusexins appearing before the radiation of eukaryotes.

Discussion

All fsx1 genes found in cultured and isolated genomes are restricted to the Haloarchaea clade. Although 83% of fsx1 genes were found in saline environments, they are not restricted to a particular niche, neither geographically nor environmentally and, by extension, potentially not restricted to halophilic archaea (Supplementary Data 1 and Fig. 6a). Only 16 out of 88 fsx1 metagenomic genes have taxonomy assigned. The 16 MAG-containing fusexins are distributed in other archaeal clades (including Asgard) and also in bacteria but we take those assignments with caution. fsx1-containing MAGs are highly fragmented and were assembled using methods that include sequence composition (k-mer) criteria, allocating scaffolds with similar k-mer spectra into MAGs. As fsx1 genes from PCGs are located within distinctive k-mer regions and the metagenomic contigs containing the fsx1 gene are compositionally homogeneous, and plausibly are also IMEs, it is quite possible that MAG-fusexins are misassigned. The association of haloarchaeal fsx1s with IMEs (Fig. 8), genomic comparisons of close species (Fig. 7a) and the incompleteness between their sequence phylogeny and cognate species tree (Fig. 6), indicate lateral mobility within the Haloarchaea class. This evidence suggests that Fsx1s mediate a cell fusion-dependent genetic exchange in archaea. This hypothesis is consistent with the genetic structure and lifestyle of halophilic archaea which are notorious for being polyploid and undergoing massive genetic exchanges that overcome species and genera barriers. Moreover, compelling evidence of archaeal cell fusion comes from studies showing bilateral DNA exchange that correlates with cytoplasmic bridges made up of fused lipid bilayers connecting haloarchaeal cells. Thus, it is plausible that Haloarchaea evolved HGT mechanisms based on conjugative-like DNA mobilization and cell-cell fusion.

The “virus or the egg” dilemma posits that fusexins may have been either viral innovations (class II fusogens), later acquired by eukaryotes, or vice versa. Archaeal fusexins expand this dilemma: gamete fusexins may have prokaryotic origins. Both structure- and sequence-based trees (Supplementary Fig. 10) do not solve but provide insights to address this conundrum, in which we distinguish three main hypotheses indicating alternative origins: Virus, Eukarya, and Archaea (Fig. 9). For all three scenarios we assume that sexual fusexins (HAP2) were present in the LECA.

A Virus-first scenario has circumstantial evidence favoring it. Extant of viral genes is documented for all three domains of life. To be consistent with the observed basal divergence between archaeal and eukaryal fusexins (Supplementary Fig. 10a) the Virus-first hypothesis must include two HGT events before the archaeal radiation (Fig. 9a). Thus, an archaeon could have exapted a fusexin from an enveloped archaeal virus and then transferred it to an early eukaryote. The reciprocal is also formally possible: an eukaryal viral fusexin was exapted by a pre-LECA eukaryotic cell and then transferred to an archaeon, before the eukaryal radiation. Alternatively, the ancestral fusexin-containing virus was able to infect both archaea and eukaryal cells (Fig. 9a). These putative events are at odds with the distribution of fusexins in extant viruses. All currently known viral fusexins belong to RNA viruses that are confined to a few multicellular hosts: vertebrates, arthropods, and flowering plants. This distribution favors a scenario where viral fusexins, like many other eukaryotic viral proteins, have eukaryotic cellular origins. Unlike currently known fusexin-containing viruses, all isolated archaeal viruses to date have DNA genomes. The recently elucidated structure of VP5, a haloarchaeal virus envelope protein that mediates cell invasion shows a fold that differs from all previously known viral fusion proteins, including fusexins. Although there is no evidence for the presence of fusexins in archaeal viruses, upcoming metagenomic and structural analyses may provide support to the Virus-first hypothesis.

The widespread presence of sexual fusexins in Eukarya indicates evolutionary success, in line with the Eukarya-first hypothesis (Fig. 9b). However, introduction of an eukaryal fusexin into the Archaeal domain is less supported by currently available evidence as interdomain gene transfers from eukaryotes to archaea are hardly documented and thought to be scarce.

The presence of fsx1 genes in Haloarchaea IMEs is consistent with gene transfer in the opposite direction. Eukaryogenesis, and by extension the emergence of sex, is marked by massive horizontal gene transfer events to the archaeal ancestor of eukaryotes, disregarding if it belonged to the Asgard superphylum or to a sister group of Archaea. During the First
Eukaryotic Common Ancestor (FECA) to LECA transition, in addition to the alphaproteobacterial endosymbiont-related inherited genes, the pre-LECA received hundreds of archaeal genes from other lineages, including Euryarchaeota\(^{54}\). Thus, an Archaea-first hypothesis (Fig. 9c) considers horizontal transfer of archaeal fusexins into the ancestor of eukaryotes. A weakness of this hypothesis is the sparse distribution of fusexins in archaeal genomes. This distribution is not a product of a recent HGT from Eukarya, as phylogenetic analysis indicates basal divergence between eukaryotic and archaeal fusexins (Supplementary Fig. 10a). Lateral mobility of extant fsx1 genes within Haloarchaea, their relative confinement to few archaeal lineages and their basal divergence from sexual fusexins suggests they are molecular relics, and that cell fusion-based HGT might have declined during archaeal evolution in favor of conjugation, transduction and transformation.

Fig. 6 Environmental distribution of Fsxl1s. a Archaeal fusexins unrooted phylogeny, environmental details, and trimeric models\(^{56}\) based on the Fsx1 X-ray structure (PDB 7P4L [https://www.rcsb.org/structure/7P4L], enlarged). Tree tip IDs coming from metagenomic data coded by sampling site, sample id and sequence (see Supplementary Data 1). Fsx1s from cultivated genomes are shown as Natrialbales and Haloferacales orders in blue and red, respectively. Surfaces colored and oriented as in Fig. 3a. Found in diverse environments and often differing in global surface properties, modeled trimers share hydrophobic tips. b Panel with phylogenetic tree for fsx1-containing cultured archaeal species, showing incongruences between Fsx1s in Haloarchaeales.
This third scenario, an Archaeal origin of fusexins, poses new challenges to both sex origins and eukaryogenesis models. Discovery of the Asgard superphylum and the recent cultivation of one of its members support eukaryogenesis scenarios where populations of bacteria and archaea lived in syntrophy, transferring metabolites and genes. Acquisition of a fusexin gene during the FECA to LECA transition could have enabled pre-LECA cells to undergo genome expansion, explore syncytial forms and evolve into mononucleated cells fully equipped for meiosis and gamete fusion. Our findings raise the possibility that gamete fusion is the product of over two billion years of evolution of this ancient archaeal cell fusion machine.

The archaeal proteins herein identified place fusexins in yet another domain of life, with different membrane chemistries and along a broad niche landscape, from cold hypersaline lakes to hot springs and hydrothermal vents (Fig. 6a). Our structural and functional analyses show that Fsx1 has both conserved and divergent properties when compared to eukaryotic and viral fusexins (Figs. 3 and 4). Like its viral counterparts, Fsx1 has an uncharged loop required for fusion. However, unlike previously known fusexins, Fsx1 harbors an additional domain (IV) involved in fusogenic activity that may bind sugars (Figs. 2d and 3d). Considering that cell surface glycosylation was found to be important for fusion-based mating of halophilic archaea, this domain may actively promote fusion by interacting with carbohydrates attached to lipids or proteins such as S-layer glycoproteins. Unlike HAP2s, Fsx1 homologs have 1–4 TMs and a variable Cys number (5–30, see Supplementary Data 1). Like eukaryotic fusexins, Fsx1 mediates BHK cell fusion in a bilateral fashion (Fig. 4d). However, in contrast to viral fusexins, the fusion activity of Fsx1 is maintained following substitution of its three TMs with a single TM or a GPI anchor. The retention of fusion activity when the transmembrane domains are replaced by a GPI anchor supports the model in which Fsx1 mediates homotypic fusion with fusogens required on both target membranes. These findings also suggest that interaction between the TMs during fusion is not essential for trimerization and expansion of the fusion pores. Since GPI-anchored AtHAP2 is also fusogenic, other fusexins may also drive complete cell fusion without a specific involvement of TMs. Future studies will address the function of the six-helix bundle formed by Fsx1 domain II, which is unprecedented among fusexins and raises an unexpected structural connection with class I viral fusexins.
What are the limitations of this study? First, the cellular and molecular functions of the fsx1 family presented in this work is reliant upon expression within a heterologous system designed to probe fusion activity. Despite membrane chemistry differing from the type that would be associated with haloarchaeal hosts of the IME, Fsx1 was able to promote membrane fusion reliably. Second, evolutionary analysis and comparative genomics confirm fsx1’s link to IMEs. These sequences appear to be spread across a wide variety of niches all across the globe despite a relatively sparse distribution in sequenced archaeal genomes. Third, in this study we focused on describing the Fsx1 family evolutionarily, structurally and functionally to place it within its context in the Fusexin superfamily, but future studies will be needed to elucidate which biological processes it is involved in as well as its relationship to the rest of the archaeal mobilome and virome. Additionally, structural features, such as the presence of a fourth domain, were modified in order to observe their effect on fusion activity within our experimental system, but it remains to be seen what importance they have in their native context. Future work will focus on studying Fsx1 in archaeal experimental systems as well as leverage metagenomic sampling and assembly techniques to exhaustively detect possible Fsx1 homologs in environments where it may be facilitating horizontal gene transfer.

**Methods**

**Initial fusexin search using structurally guided MSAs**. HMMs were prepared using structurally guided multiple sequence alignments (MSAs) of known eukaryotic HAP2 sequences (ectodomains only). Structural MSAs were derived using structurally guided multiple sequence alignments (MSAs) of known eukaryotic fusexin homologs, covering major lineages, was manually curated. A MSA was built for each homolog by using the sequence as a query on the Uniclust database (nr90) with HHblits for three iterations. This set of MSAs was compiled into an HH-suite database and each MSA was used as a query against this database to establish a profile-based distance matrix using the probability of homology (Fig. 1a and Supplementary Fig. 1).

**HMM-based distance matrices**. A taxonomically representative list of known viral and eukaryotic fusexin homologs, covering major lineages, was manually curated. A MSA was built for each homolog by using the sequence as a query on the Uniclust database with HHblits for three iterations. This set of MSAs was compiled into an HH-suite database and each MSA was used as a query against this database to establish a profile-based distance matrix using the probability of homology (Fig. 1a and Supplementary Fig. 1).

**Metaclust database search pipeline**. We searched the Metaclust dataset (nr50) using an HMM made of Fsx1 sequences found in PCGs and MAGs (Supplementary Data 1; see also codes, notebooks and datasets available at Zenodo). Fsx1 sequences were aligned using ClustalO with default settings for 3 iterations and the resulting MSA was used as a query with HHMER hmmssearch against the Metaclust50 dataset. All returned sequences with an E-value < 0.0001 and an alignment length greater than 100 residues were ranked for further analysis. PSI-BLAST was also used on the Metaclust (nr90) with Fsx1 sequences found in PCGs and MAGs with default parameters for 3 iterations. All returned sequences with an E-value < 0.0001 and an alignment length greater than 100 were added to the pool of candidates. Manual curation was performed using membrane protein topology predictor TOPCONS and distant homology searches using HHblits against PDB70.

**DNA constructs**. Ten archaean genes were synthesized (GenScript) and cloned into pGene/V5-His vectors (Supplementary Table 5). Details of nucleotides used for synthesis and protein sequences are described in Supplementary Data 4. For structural studies, a synthetic gene fragment encoding the extracellular region of a metagenomic Fsx1 ORF (IMG genome 3300800868, scaffold JGI12330J12834_1000008, ORF 8; Supplementary Data 1) was subcloned by PCR in frame with the 5’ chicken Crypa signal peptide and 3’ 8xHis-tag-encoding sequences of pLJ6, a mammalian expression vector derived from Chlamydomonas reinhardtii (PDB 6E18 [https://www.rcsb.org/structure/6E18]) and Arabidopsis thaliana (PDB 5OW3 [https://www.rcsb.org/structure/5OW3]).
AtHAP2-V522, EFF-1-V5, VSV-G15 and other archaeal fusexins (NaFsx1, pCI::H2B-GFP plasmid (see list of primers in Supplementary Table 7). Fsx1-V5, BmgBI and BglII and used to replace the H2B-GFP coding sequence of the presence of fsx1 Table 6), an oligo DNA encoding for the nuclear export signal (LQKKLEELELD) constructs were veri
included residues D25-S535 and contained a T369C substitution, introduced by details are found in Supplementary Tables 6 and 7.

To generate pCI-GFPnes plasmid (see list of plasmids in Supplementary Table 6), an oligo DNA encoding for the nuclear export signal (LQKKLEELELD) was cloned downstream the region encoding EGFP of the pCAGG plasmid using the enzyme BsrGI. Then, the GFPnes coding sequence was amplified, cut with BnmBII and BglII and used to replace the H2B-GFP coding sequence of the pC:HH2B-GFP plasmid (see list of primers in Supplementary Table 7). Fsx1-V5, AtHAP2-V535, EFF-1-V5, VSV-G15 and other archaeal fusexins (NaFsx1, HQ22Fsx1, HnFsx1) were subcloned into corresponding pCI::H2B-RFP/H2B-GFP/GFPnes vectors separately. For mutagenesis of Fsx1, (i) Fsx1-DIV-EFF-1 stem: The stem region of EFF-1 (E510-D561) was amplified from pGene::EFF-1-V5 and fused to the upstream and downstream regions of Fsx1-DIV with overlapping primers. (ii) Fsx1-DIV-EFF-1: The TM and cytoplasmic tail of EFF-1 (I562–I658) were amplified from pGene::EFF-1-V5 and fused to the ectodomain of Fsx1 to replace its TMs. (iii) Fsx1ΔTMs→EFF-1 TM: The TM and cytoplasmic tail of EFF-1 (I562–I658) were amplified from pGene::EFF-1-V5 and fused to the ectodomain of Fsx1 to replace its TMs.

Protein expression and purification. HEK293T cells (ATCC CRL-3216) were transiently transfected using 25 kDa branched polyethyleneimine and cultured in DMEM media (Invitrogen) supplemented with 2% (v/v) fetal bovine serum (Biological Industries). 90 h after transfection, the conditioned media from HEK293T cells was harvested, 0.2 μm-filtered (Pall) and adjusted to 20 mM Na-HEPES pH 7.8, 2.5 mM NaCl, 35 mM imidazole. 10 mM Ni Sepharose excel beads (GE Healthcare) pre-equilibrated with immobilized metal affinity chromatography (IMAC) buffer (20 mM Na-HEPES pH 7.8, 2.5 mM NaCl, 10 mM imidazole) were added to 11 adjusted conditioned media and incubated overnight at 4 °C. After washing the beads with 100 column volumes IMAC buffer, captured Fsx1E was batch-eluted with 30 mL 20 mM Na-HEPES pH 7.8, 2.5 mM NaCl, 500 mM imidazole, and concentrated with 30 kDa-cutoff centrifugal filtration devices (Amicon). The material was then further purified by SEC at 4 °C, using an AKTAplc chromatography system (GE Healthcare) equipped with a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM Na-HEPES pH 7.8, 2.5 mM NaCl. Peak fractions were pooled and concentrated to 5 mg mL−1 (Supplementary Fig. 2a, b).

Size-exclusion chromatography-multiangle light scattering (SEC-MALS). Purified Fsx1E samples (120–150 μg) were measured using an Elan LC high-performance liquid chromatography system with a UV-900 detector (Amerham Pharmacia Biotech; λ = 280 nm), coupled with a miniDawn TRES MALS detector (Wyatt Technology; λ = 685 nm) and an Optilab T-rEX dRI detector (Wyatt Technology; λ = 660 nm). Separation was performed at 20 °C using a Superdex 200 Increase 10/300 GL column (GE Healthcare) with a flow rate of 0.5 mL min−1 and mobile phases consisting of 20 mM Tris–HCl pH 8.5, 150 mM NaCl (normal salt condition) or 20 mM Tris–HCl pH 8.5, 2.0 mM NaCl and 0.2 M CaCl₂ (high salt condition) (Supplementary Fig. 2c). Data processing and weight-averaged molecular mass calculations were performed using ASTRA (Wyatt Technology). BSA (150 μg) was used as a control.

Small-angle X-ray scattering (SAXS). SAXS experiments were performed at beamline BM29 of the European Synchrotron Radiation Facility (ESRF), using Fsx1E (4.5 mg mL−1) in 20 mM Na-HEPES pH 7.8, 150 mM NaCl. Sample delivery and measurements were performed using a 1 mm-thick quartz capillary, which is part of the BM29 BioSAXS automated sample changer unit81. Data were collected at 1 Å wavelength in 10 frames of 1 s at 20 °C, using an estimated beam size of 1 mm × 100 μm; buffer blank measurements were carried out under the same conditions, both before and after sample measurement. Data were averaged and subtracted using PRIMUS82 from the ATSAS package83, which was also used to calculate the pair-distance distribution function, as well as the radius of gyration and the Porod volume. Theoretical scattering curves for monomeric and trimeric Fsx1E were calculated and compared with the experimental data using CRYOSOL84. Ab initio envelope reconstruction was performed with DAMMIF85, resulting in 20 models that were superimposed and averaged with DAVANA86. Chain A of the refined Fsx1E model was either rigidly fitted with UCSF ChimeraX87 into the
envelope generated by averaging all 20 independent ab initio SAXS models (Supplementary Fig. 2d, top envelope), or flexibly fit with Namdinator to the average envelope of the two templates of the three clusters of SAXS models (accounting for 4 and 5 of the 20 SAXS models, respectively; Supplementary Fig. 2d, middle and bottom envelopes).

Crystallographic and X-ray diffraction data. Two similar initial hits obtained from extensive screening using a mosquito crystallization robot (TTP Labtech) were manually optimized by setting up vapor diffusion experiments at 20°C in 24-well plates. To grow diffraction quality crystals, 1 µl concentrated Fsx1 was mixed with 1 µl 23% (w/v) PEG 4000, 0.1 M Tris–HCl pH 8.5, 0.2 M CaCl2, and equilibrated against 1 ml of the same solution. Romboidal plates of Fsx1 grew in 1–3 months from protein precipitate that appeared after overnight equilibration of the crystal growth drops (Supplementary Fig. 2e). For data collection, specimens were frozen from thin coverslip by cryomicrotomed with Micromanipulator (MiTeGen) and flash frozen in liquid nitrogen. More than a hundred crystals were screened at beamlines ID32-1 of the ESRF and I04 of Diamond Light Source, yielding datasets of highly variable quality. The final X-ray diffraction dataset at 2.3 Å resolution was collected at ESRF ID32-1.

Data reduction and non-crystallographic symmetry analysis. Datasets were processed in space group C2 with XDS (Supplementary Table 2). By revealing a strong non-origin peak at chi = 120 (Supplementary Fig. 2f), self rotation functions calculated with MOLREP91 or POLARIN92 clearly indicated the presence of three-fold non-crystallographic symmetry (NCS) within the asymmetric unit of the centered monomeric crystals. Combined with Matthews coefficient calculations93,94, this strongly suggested that Fsx1 crystalized as a homotrimer.

Structure determination by molecular replacement with AlphaFold2 models. Multiple attempts to experimentally determine the structure of Fsx1 using a variety of heavy atoms failed, probably because the high-salt mother liquor composition hindered heavy atom binding. Because molecular replacement (MR) with HAP2-derived homology models also failed, we phased the data by taking advantage of the recent significant advances in protein 3D structure prediction using machine learning95 to phase the data by MR96 (Supplementary Fig. 3). For data collection, specimens were frozen from thin coverslip by cryomicrotomed with Micromanipulator (MiTeGen) and flash frozen in liquid nitrogen. More than a hundred crystals were screened at beamlines ID32-1 of the ESRF and I04 of Diamond Light Source, yielding datasets of highly variable quality. The final X-ray diffraction dataset at 2.3 Å resolution was collected at ESRF ID32-1.

Sequence-structure analysis. Sequence alignments were rendered with ESPript97 and manually annotated. Transmembrane helices were predicted using TMHMM98. GDT TS scores were calculated using LGA100 and structural similarities were assessed with Dal111 and PDBeFold102. Secondary structure was assigned using DSSP111. Subunit interfaces were analyzed by PDBsum114, Pic115 and PDBePISA116. Molecular charge was calculated using the YASAR2 force field of YASARA Structure117 and electronic surface potential calculations were performed with PDBePQR118 and APBS119, via the APBS Tools plugin of PyMOL (Schrodinger, LLC). Mapping of amino acid conservation onto the 3D structure of Fsx1E was carried out by analyzing a sequence alignment of archaeal homologs with ConsSurf120. Structural figures were generated with PyMOL.

Structural modeling of trimeric Fsx1. Models of homotrimeric Fsx1 were generated using a local copy of AlphaFold-Multimer121, installed using the open source code and instructions available at https://github.com/deepmind/alphafold.

Cells and reagents. Baby Hamster Kidney (BHK-21) cells (kindly obtained from Judith White, University of Virginia) were maintained in DMEM supplemented with 10% FBS (Biological Industries), 100 U/ml penicillin, 100 µg/ml streptomycin (Biological Industries), 2 mm l-glutamine (Biological Industries), 1 mM sodium pyruvate ( Gibico), and 30 mM HEPES buffer, pH 7.3, at 37°C with 5% CO2. Transfections were performed using Fugene HD (Promega) or jetPRIME (Polyplus) according to the manufacturer’s instructions.

Immunofluorescence. BHK cells were grown on 24-well tissue-culture plates with glass coverslips. Permeabilized cells were fixed with 4% paraformaldehyde (EM grade, Bar Nao, Israel) in PBS, followed by incubation in 40 mM NH4Cl to block free aldehydes, permeabilized in 0.1% Triton X-100 in PBS and blocked in 1% FBS in PBS for 5 min. After fixation, the coverslips were incubated 1 h with mouse anti-V5 antibody (Invitrogen, 1:500) and 1 h with the secondary antibody which was donkey anti-mouse coupled to Alexa Fluor 488 (Invitrogen, 1:50). Alternatively, for immunofluorescence without permeabilization, cells were blocked on ice in PBS with 1% FBS for 20 min, and then stained with Monoclonal ANTI-FLAG M2 antibody (Sigma, 1:1000) on ice for 1 h. After anti-FLAG staining, cells were washed and fixed with 4% PFA in PBS. Cells were blocked again and stained with the secondary antibody (donkey anti-mouse coupled to Alexa Fluor 488; Invitrogen diluted 1:500 in PBS for 1 h. In all cases, nuclei were stained with 1 µg/ml DAPI. Images were captured using a Nikon Eclipse E800 with a 60x/1.40 Plan ApoNomarsk objective and a Nikon C2+ high-sensitivity EMCCD camera (Nikon) at 300° K and a Digital Micrograph ORCA-ER camera controlled by Micro-Manager software122 (Fig. 5).

Western blots. 24 h post-transfection, cells were treated with lysis buffer (50 mM Tris–HCl pH 8.0, 100 mM NaCl, 5 mM EDTA, 1% Triton X-100 supplemented with chymostatin, leupeptin, antipain and pepstatin) on ice for 10 min. After centrifugation at 21,000 × g for 10 min at 4°C, supernatants of lysates were mixed with reducing sample buffer (+DTT) and incubated 5 min at 95°C. Samples were loaded on a 10% SDS-PAGE gel and transferred to PVDF membrane. After blocking, membranes were incubated with primary antibody anti-V5 mouse monoclonal antibody (1:5000; Invitrogen) or anti-actin (1:2000; MP Biomedicalals) at 4°C overnight and HRP-conjugated goat anti-mouse secondary antibody 1 h at room temperature. Membranes were imaged by the ECL detection system using FUSION-PULSE E6 (VILBER).
ARTICLE

pCE-V5-Cc:GFpines in respective 35 mm plates. The cells were incubated, washed, and mixed with pCL-H2B-RFP (empty vector) transfected cells (Fig. 4e). For evaluating the mutants, BHK-21 cells were transfected with 1 µg pCI-Fsx1-V5:GFpines or pCI-Fsx1-V5:H2B-RFP or the plasmids encoding for each mutant: AFl → AGA, ΔDIV → EFF-1 stem, ATMs → EFF-1 TM, Fxs1ATMs → GPl, or AtfHAP2ATMs → GPl (Fig. 5a). Empty pCI-GFPines or pCL-H2B-RFP were used as negative controls. 4 h after transfection, the cells were washed, counted, mixed, and incubated as previously described. In all cases, 18 h after mixing, 20 µM FdUrd was added to the plates, and 24 h later, the cells were fixed with 4% paraformaldehyde diluted in PBS. Nuclei were stained with 1 µg/ml DAPI. Images were obtained using wide-field illumination with an ELYTRA system S1 microscope as described above.

The GFP + RFP mixing index was calculated as the number of nuclei in mixed cells, green cytoplasm (GFpines) with red (H2B-RFP) and blue (DAP) nuclei out of the total number of nuclei in fluorescent cells in contact (Figs. 4f, 5b). The multinucleus index was defined as the ratio between the number of nuclei in multinucleated cells (N1) and the total number of nuclei in multinucleated cells and expressing cells that were in contact (N2) but did not fuse, using the following equation: % multinucleation = N1(N1 + N2) × 100. The percentage of multinucleation was calculated for GFpines cells with RFP and DAPI nuclei. For the unilateral assay, multinucleation was determined as the ratio between the number of nuclei in multinucleated green cells and the total number of nuclei in green multinucleated cells and GFpines expressing cells that were in contact but did not fuse (Fig. 4f).

**Live imaging of fusing cells.** BHK cells were plated on 15 mm glass bottom plates (Wuxi NEST Biotechnology Co., Ltd.) and transfected with 1 µg pCI-Fsx1-V5:H2B-GFP together with 0.5 µg myristoylated-mCherry (myr-palm-mCherry; kindly provided by Valentine Danusing and Salvatore Chiantia12). 18 h after transfection, the cells were washed with 2 µg/ml methylene dye for 10 min at 37 °C and washed once with fresh medium. Time lapse microscopy to identify fusing cells was performed using a spinning disk confocal microscope (CSU-X1; Yokogawa Electric Corporation) with an Eclipse Ti and a Plan-Apochromat ×20 (NA, 0.75; Nikon) objective. Images in differential interference contrast and red channels were recorded every 4 min in different positions of the plate using high gain and minimum laser exposure. Time lapse images were captured with an Ixon 3 EMCCD camera (Andor Technology). After 5 h, confocal z-series, including detection of the DAPI channel, were obtained to confirm the formation of multinucleated cells. Image analyses were performed in MetaMorph (Molecular Devices) and ImageJ124 (National Institutes of Health).

**Surface biotinylation.** Proteins localizing on the surface were detected as previously described12. Briefly, BHK cells were transfected with 1 µg pCAGGS, pCAGGS-EF-1 V5, pCAGGS:Fxs1-V5, pCAGGS:AFl → AGA-V5, pCAGGS:ΔDIV → EFF-1 stem-V5 or pCAGGS:Fxs1ATMs → EFF-1 TM-V5. 24 h later, the cells were washed with ice-cold PBS2 + (Ca2+ and Mg2+) and incubated with 0.5 µg/ml EZ-Link Sulfo NHS-Biotin (Thermo Fisher Scientific) for 30 min on ice. The cells were washed four times with ice-cold PBS2 +, once with DMEM with 10% FBS (to quench residual biotin), washed with PBS2 −, and resuspended in 100 µl 0.3% SDS. After an incubation of 12 h at 4 °C the resin was separated by centrifugation (2 min at 21,000 × g) and the lysate was mixed with NeutrAvidin Agarose Resin (Thermo Fisher Scientific) and washed once with fresh medium. Time-lapse microscopy to identify fusing cells was performed using a spinning disc confocal microscope (CSU-X1; Yokogawa Electric Corporation) with an Eclipse Ti and a Plan-Apochromat ×20 (NA, 0.75; Nikon) objective. Images in differential interference contrast and red channels were recorded every 4 min in different positions of the plate using high gain and minimum laser exposure. Time lapse images were captured with an Ixon 3 EMCCD camera (Andor Technology). After 5 h, confocal z-series, including detection of the DAPI channel, were obtained to confirm the formation of multinucleated cells. Image analyses were performed in MetaMorph (Molecular Devices) and ImageJ124 (National Institutes of Health).

**Integrated mobile element (IME) identification by k-mer spectra analysis and comparative genomics.** Comparison between close species with presence (fsx1+ or absence (fsx1−) of aerialich fungal xenases to detect insertion sites was done by performing sequence similarity searches in complete genomes from the closest relatives available in the PATRIC database12 (Fig. 7a and Supplementary Table 4). Coordinates of fsx1-containing IMEs present in PCGs are annotated in Supplementary Table 4.

Among different methodologies that rely on DNA composition to identify horizontally transferred genomic regions12, k-mer spectrum analysis is a standard tool for this purpose12,12. Normalized k-mer spectra for DNA sequences of arbitrary length were generated by counting occurrences of all k-mer sizes and normalizing by the total amount of words counted. k-mer sizes from 3 to 20 bp were tested with no effect on results. A length of 4 bp was selected. To detect possible horizontally transferred regions, an average spectrum for each genome was calculated. A spectrum was calculated for a sliding window of 1 kb using 500 bp steps and subtracted from the genomic average at each window position (Fig. 7b). The absolute difference between the genome average and window spectra is represented over the entire genome. Gaussian mixture models using two distributions were fitted12 to the k-mer content of all windows, to classify these as belonging to either the core genome or transferred elements. This deviation in k-mer spectra has been explored in the context of the archaeal mouldome and contains information on the ecological niche and evolutionary history of DNA sequences10.

**IME gene content and homology analyses.** We followed the pipeline depicted in Supplementary Fig. 8. Briefly, PCGs’ IMEs were determined by a combination of k-mer spectra and genomic alignments (see Supplementary Table 4). We initially inspected each fsx1-containing scaffold to confirm that only sequence derived from an IME was found for longer or downstream analyses. We generated an enriched annotation for each IME. Then, we obtained an initial set of groups of homologous sequences, and each of these groups was enriched by means of HMM searches. Subsequently, the enriched homology groups showing similarity between them, as judged by HMM-EH homology comparisons, were merged.

In detail, first, we re-annotated the identified mobile elements combining the corresponding segment of the PATRIC125 (GFF annotation file with in-house ORF predictions) (minimum ORF length of 30 nucleotides, option by default). ORF inference was done by means of getorf of the EMBOSs package13, specifying genetic code by Table 11 (Bacteria and Archaea) and other parameters running by default. The similarity of inferred ORFs and annotated features in these mobile elements (i.e., features in their GFF annotation file) was established by means of BLASTP reciprocal searches14. We kept all the predicted ORFs and homologs that were at least annotated in one genome, in this way we tried to recover misannotated conserved ORFs.

Initial sets of homologs were generated with get_homologs12. Sequence identity and query coverage thresholds were set to 35% and 70%, respectively. In paralogues were not allowed within these groups (option ‘-e’), and remaining parameters were run by default. Homolog profiles were constructed for each homology group. To this aim, homologous sequences were retrieved for members of each group from the UniRef50 database13 with jackhammer from the HMMER package73 running with one iteration (‘-N 1’ parameter). MSAs were then generated for each group and its relevant hits with MAFFT134 running under ‘auto’ parameter, and HMMVirus were created with HMMER. Homolog groups were enriched by means of HMM searches with HMMER hmmscan, using each HMM as a query against a database comprising all predicted ORFs described above. Hits showing an e-value < 1e−10 and covering at least 50% of the HMM were added to the groups.

Enriched homology groups showing homology were collapsed. For this purpose, HMM-vs-HMM comparisons were performed with HHblits of the HHsuite135. A graph was created with the networks Python library (https://networks.org), each node being an enriched group of homologs. An edge was established between nodes if their HMM-HMM alignment was significant (i.e., e-value < 1e−10, HMM coverage of > 95%, identities of > 50%). Groups of inter-connected nodes were established with the ‘connected_components’ routine, creating a collapsed homology group in each case. Finally, we assessed the gene content similarity between mobile elements using a Jaccard Index based on the homology groups defined above. Usual Jaccard index of two sets is defined as (# of the intersection)/(# of the union). In this case:

\[ J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]

**Number of homology groups shared between MEs as a percentage of total (J/A, B)\%

We performed a hierarchical clustering of the MEs based on a distance matrix obtained from the pairwise Jaccard Indexes (distance(A,B) = 1 − J(A,B)). This was done in Python with seaborn136, employing the clustermap function. A subset of 11 mobile elements (red cluster in Supplementary Fig. 9), which included ME from PCGs and [GI:121330]128431000008 (Supplementary Data 1, 3 and 4), was selected for synteny conservation analysis. Plots depicting synteny in gene content between homolog groups were generated employing the MCscan tool137. HMMER and Pfam138 were used on default parameters to assign domains and their associated aroCOG139,140 identifiers to ORFs (Supplementary Data 2).

These analyses, including collapsed clusters, can be found in Zenodo136.
Received: 27 May 2022; Accepted: 22 June 2022; Published online: 06 July 2022

References

1. Speijer, D., Lukej, J. & Eliáš, M. Sex is a ubiquitous, ancient, and inherent feature of eukaryotic life. Proc. Natl Acad. Sci. USA 112, 8827–8834 (2015).
2. Ramesh, M. A., Malik, S.-B. & Logsdon, J. M. Jr. A phylogenomic inventory of synteny analyses (Supplementary Figs. 8 and 9, Fig. 8 and Supplementary Data 2 and 3); and structure- and sequence-based comparisons (Supplementary Fig. 10 and Supplementary Data 5), are provided on GitHub (https://github.com/DessimozLab/Archaeo-Fusexins) and Zenodo (https://doi.org/10.5281/zenodo.6677729)(5).

Code availability
All relevant codes, notebooks and datasets necessary for: HHblits and HMMER searches and structure factors and atomic coordinates have been deposited in the Protein Data Bank under accession code 7P4L. Source Data for this paper’s structure-function analyses (Figs. 4 and 5, and Supplementary Fig. 1) are provided in the Source Data file; sequences of synthesized fsx1 genes are in Supplementary Data 4. Source data are provided with this paper.

for common mechanisms of sexual reproduction shared by plants and protozoa and for male-female differentiation. Dev. Biol. 415, 6–13 (2016).
2. Valanassi, C. et al. Arabidopsis HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens. J. Cell Biol. 216, 571–581 (2018).
3. Fédry, J. et al. The ancient gamete fusogen HAP2 is a Eukaryotic Class II fusion protein. Cell 168, 904–915.e10 (2017).
4. Pinello, J. F. et al. Structure–function studies link Class II viral fusogens with structural and sequence-based comparisons. Nat. Commun. 7, 651–660 (2017).
5. Fédry, J. et al. Evolutionary diversification of the HAP2 membrane insertion motifs to drive gamete fusion across eukaryotes. PLoS Biol. 16, e2006357 (2018).
6. Feng, J. et al. Fusion surface structure, function, and dynamics of gamete fusogen HAP2. Elife 7, e39772 (2018).
7. Wongs, J. L. & Johnson, M. A. Is HAP2-GCS1 an ancestral gamete fusogen? Trends Cell Biol. 20, 134–141 (2010).
8. Dom, R. W. What came first-the virus or the egg? Cell 168, 755–757 (2017).
9. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
10. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. Structure of the dengue virus envelope protein after membrane fusion. Nature 427, 313–319 (2004).
11. Mueller, G. A. et al. Serological, genomic and structural analyses of the major mite allergen Der p 23. Clin. Exp. Allergy 46, 365–376 (2016).
12. Tunyavavunakool, K. et al. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596 (2021).
13. DeBois, R. M. et al. Functional and evolutionary insight from the crystal structure of rubella virus protein E1. Nature 493, 552–556 (2013).
14. Dubé, M., Etienne, L., Fels, M. & Kielan, M. Calcium-dependent Rubella virus fusion occurs in early endosomes. J. Virol. 90, 6303–6313 (2016).
15. Gattegno, T. et al. Genetic control of fusion pore expansion in the epidermis of Caenorhabditis elegans. Mol. Biol. Cell 18, 1153–1166 (2007).
16. Kemble, G. W., Daniell, T. & White, J. M. Lipid-anchored hemagglutinin promotes hemifusion at complete fusion. Cell 76, 383–391 (1994).
17. Markosyan, R. M., Cohen, F. S. & Melikyan, G. B. The lipid-anchored ectodomain of influenza virus hemagglutinin (GPI-HA) is capable of inducing nonenveloped fusion pores. Mol. Biol. Cell 11, 1143–1152 (2000).
18. Jones, N. A. & Geraghty, R. J. Fusion activity of lipid-anchored envelope glycoproteins of herpes simplex virus type 1. Virology 324, 213–228 (2004).
19. Lutz, K. & Soppa, J. Polyploidy in halophytic archaea: regulation, evolutionary advantages, and gene conversion. Biochem. Soc. Trans. 47, 933–944 (2019).
20. Turgeman-Grott, I. et al. Pervasive acquisition of CRISPR memory driven by inter-species mating of archaea can limit gene transfer and influence speciation. Nat. Microbiol. 4, 177–186 (2019).
21. DeMaere, M. Z. et al. High level of intergeneric gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. Proc. Natl Acad. Sci. USA 110, 16939–16944 (2013).
22. Sivalasundra, S. et al. Analysis of cell-cell bridges in Halofexax volcanii using electron cryo-tomography reveals a continuous cytoplasm and S-layer. Front. Microbiol. 11, 61239 (2020).
23. Rosenshine, L., Tchelet, R. & Mevarech, M. The mechanism of DNA transfer in the mating system of an archaeabacterium. Science 245, 1387–1389 (1989).
24. Naor, A., Lapierre, P., Mevarech, M., Papke, R. T. & Gophna, U. Low species barriers in halophilic archaea and the formation of recombinant hybrids. Curr. Biol. 22, 1444–1448 (2012).
25. Wagner, A. et al. Mechanisms of gene flow in archaea. Nat. Rev. Microbiol. 15, 492–501 (2017).
26. Koonin, E. V. & Krupovic, M. The depths of virus extinction. Curr. Opin. Virol. 31, 1–8 (2018).
27. Modis, Y. Relating structure to evolution in class II viral membrane fusion proteins. Curr. Opin. Virol. 5, 34–41 (2014).
28. Merchant, M. et al. A bioactive phlebovirus-like envelope protein in a hokkowendogenous virus. Sci. Adv. 8, eabf6994 (2022).
29. Krupovic, M. & Koonin, E. V. Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl Acad. Sci. USA 114, E2401–E2410 (2017).
30. El Omari, K. et al. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. Nat. Commun. 10, 846 (2019).
31. Brochier-Armanet, C. et al. Complete-fosmid and fosmid-end sequences reveal frequent horizontal gene transfers in marine uncultured planktonic archaea. ISME J. 5, 1291–1302 (2011).
32. Gábaldón, T. Origin and early evolution of the Eukaryotic cell. Annu. Rev. Microbiol. 75, 631–647 (2021).
33. Liu, Y. et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature 593, 553–557 (2021).
34. Bosseberg, J. et al. Timing the origin of eukaryotic cellular complexity with ancient duplications. Nature 593, 59–64 (2021).
35. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353–358 (2017).
Author contributions
B.P. conceived the experiments; performed some imaging work; designed, supervised and analyzed cell fusion experiments. C.V. designed, performed and analyzed cell fusion assays. DdeS and K.T. made constructs of Fsx1E mutants. K.F. made constructs of Fsx1 mutants. K.T. and L.J. performed imaging work. N.G.B., M.G., H.R., L.J., and B.P. made fluorescent protein tags for quantifying protein oligomerization in living cells. Sci. Rep. 8, 1–12 (2018).

Acknowledgements
We thank Sonja-Yerena Albers, Dan Cassel, Alejandro Colman-Lerner, Uri Gophna, Yael Issawi-Sklevitz, Shahar Lavid, Peter Walter, as well as members of our laboratories for discussion and comments on the manuscript; Jose Flores for advice on searches of fusinex in metagenomes; Olivier Gascuel for discussions on phylogenetics; Yoav Henis for the GPl-BHA plasmid; Kira Makarova for sharing insights on mobile elements. Part of the computations for this work was performed at the Vital IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. This work was supported by Beca de Doctorado Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina (D.M.); Comisión Sectorial de Investigación Científica grant CSIC 1+D-2020-682, Uruguay (H.R., M.G.); European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 848407 (N.G.B.); FOCEM, Fondo para la Convergencia Estructural del Mercosur grant COF 03/11 (M.G.); Fondo para la Investigación Científica y Tecnológica grant, Argentina (PICT-2017-0854 (P.S.A.); Israel Science Foundation grants 257/17, 2462/18, 2327/19, and 178/20 (B.P.); Knut and Alice Wallenberg Foundation grant 2018:0042, Sweden (L.J.); Novartis Foundation for Medical-Biological Research grant 17B11, Switzerland (C.D.); Programa de Desarrollo de las Ciencias Básicas, Uruguay (H.R., M.G., and M.L.); Swedish Research Council grants 2016-03589 and 2020-03496 (L.J.); Swiss Leading House for the Latin American Region (C.D. and P.S.A.); Swiss National Science Foundation grant 183723 (C.D.).

Author contributions
B.P. conceived the experiments; performed some imaging work; designed, supervised and analyzed cell fusion experiments. C.V. helped devise analysis strategies for k-mer and phylogenetic analysis. C.V. designed, performed and analyzed cell fusion assays. DdeS collected X-ray data, took part in structural analysis and validated metal substructure. D.M. carried out deep homology detection of Fsx1; designed gene context analysis strategies and phylogenetic analysis, collected sequence data; performed k-mer, functional, structural and phylogenomics analyses, built homology models; coded analysis routine pipelines. H.R. supervised the bioinformatics part of the work; designed and performed sequence phylogenetic analyses and phylogenomic surveys. K.F. made constructs of Fsx1 mutants. K.T. and L.J. generated AlphaFold models of monomeric Fsx1E. L.J. supervised the biochemical and structural part of the work; collected X-ray data; solved the Fsx1 structure, refined and analyzed it; generated AlphaFold-Multimer models of trimeric Fsx1. M.G. designed bioinformatic strategies; supervised bioinformatic aspects of the work; analyzed sequence and structural data. M.L. performed I.M.E. synteny analyses, phylogenetic and phylogenomic surveys. N.G.B. carried out live imaging and surface biotinylation experiments; assisted with the preparation and design of plasmids; analyzed data; P.S.A. supervised the bioinformatics part of the work and analyzed data. S.N. expressed, purified and crystallized Fsx1; performed SEC-MALS experiments; analyzed SAXS data; collected X-ray data and took part in structure determination, model building and structure analysis. X.L. designed and constructed plasmids; carried out immunofluorescence and western blots for archaean fusinex in mammalian cells; performed imaging work. D.M., S.N., X.L., N.G.B., M.G., H.R., P.S.A., L.J. and B.P. made figures and tables. D.M., S.N., X.L., M.G., H.R., P.S.A., L.J. and B.P. wrote the manuscript, which was reviewed by all authors.

Competing interests
J.L. has filed provisional patent applications relating to machine learning for predicting protein structures. The other authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-31564-1.

Correspondence and requests for materials should be addressed to Martin Graña, Héctor Romero, Pablo S. Aguilar, Luca Jovine or Benjamin Podbilewicz.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
