



# The structures of two archaeal type IV pili illuminate evolutionary relationships

Fengbin Wang, Diana P Baquero, Zhangli Su, Leticia C Beltran, David Prangishvili, Mart Krupovic, Edward Egelman

## ► To cite this version:

Fengbin Wang, Diana P Baquero, Zhangli Su, Leticia C Beltran, David Prangishvili, et al.. The structures of two archaeal type IV pili illuminate evolutionary relationships. Nature Communications, 2020, 11 (1), pp.3424. 10.1038/s41467-020-17268-4 . pasteur-02894969

**HAL Id: pasteur-02894969**

**<https://pasteur.hal.science/pasteur-02894969>**

Submitted on 9 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ARTICLE



<https://doi.org/10.1038/s41467-020-17268-4>

OPEN

# The structures of two archaeal type IV pili illuminate evolutionary relationships

Fengbin Wang<sup>1,5</sup>, Diana P. Baquero <sup>2,3,5</sup>, Zhangli Su<sup>1</sup>, Leticia C. Beltran<sup>1</sup>, David Prangishvili<sup>2,4</sup>, Mart Krupovic<sup>2</sup>  & Edward H. Egelman <sup>1</sup> 

We have determined the cryo-electron microscopic (cryo-EM) structures of two archaeal type IV pili (T4P), from *Pyrobaculum arsenaticum* and *Saccharolobus solfataricus*, at 3.8 Å and 3.4 Å resolution, respectively. This triples the number of high resolution archaeal T4P structures, and allows us to pinpoint the evolutionary divergence of bacterial T4P, archaeal T4P and archaeal flagellar filaments. We suggest that extensive glycosylation previously observed in T4P of *Sulfolobus islandicus* is a response to an acidic environment, as at even higher temperatures in a neutral environment much less glycosylation is present for *Pyrobaculum* than for *Sulfolobus* and *Saccharolobus* pili. Consequently, the *Pyrobaculum* filaments do not display the remarkable stability of the *Sulfolobus* filaments in vitro. We identify the *Saccharolobus* and *Pyrobaculum* T4P as host receptors recognized by rudivirus SSRV1 and tristromavirus PFV2, respectively. Our results illuminate the evolutionary relationships among bacterial and archaeal T4P filaments and provide insights into archaeal virus-host interactions.

<sup>1</sup>Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA. <sup>2</sup>Archaeal Virology Unit, Department of Microbiology Institut Pasteur, 25 rue du Dr. Roux, Paris 75015, France. <sup>3</sup>Sorbonne Universités, Collège Doctoral, Paris 75005, France. <sup>4</sup>Ivane Javakhishvili Tbilisi State University, Tbilisi 0179, Georgia. <sup>5</sup>These authors contributed equally: Fengbin Wang, Diana P. Baquero. ✉email: [mart.krupovic@pasteur.fr](mailto:mart.krupovic@pasteur.fr); [egelman@virginia.edu](mailto:egelman@virginia.edu)

It has been recognized for many years that bacterial type IV pili (T4P)<sup>1,2</sup>, important in activities such as adhesion<sup>3,4</sup>, twitching motility<sup>5</sup>, surface-sensing<sup>6</sup> and natural transformation<sup>7</sup>, are homologs of two different classes of archaeal surface structure, pili and flagellar filaments<sup>8–11</sup>. The archaeal flagellar filament is a unique rotating T4P-like assembly which enables motility and is conserved in widely different archaeal lineages, including hyperthermophiles, halophiles and methanogens<sup>12</sup>. Archaeal T4P are even more widespread and functionally diverse<sup>13</sup>. Although the full spectrum of functions played by archaeal T4P is yet to be uncovered, they are known to mediate adhesion to various biotic and abiotic surfaces, and play an important role in DNA exchange, intercellular communication and biofilm formation<sup>14,15</sup>. Furthermore, T4P serve as receptors for certain hyperthermophilic archaeal viruses, including *Sulfolobus islandicus* rod-shaped viruses 2 and 8 (SIRV2 and SIRV8, respectively; family *Rudiviridae*)<sup>16,17</sup> and *Sulfolobus turreted* icosahedral virus (STIV; family *Turriviridae*)<sup>18</sup>.

Despite the importance of bacterial and archaeal T4P-like filaments, the structural details of these filaments remain sparse. However, the revolution in cryo-EM has now led to atomic structures for seven bacterial T4P<sup>19–21</sup>, three archaeal flagellar filaments<sup>22–24</sup>, one archaeal “flagellar-like” filament<sup>25</sup>, and one archaeal T4P<sup>26</sup>. Interestingly, a 4.1-Å resolution structure of the T4P from *S. islandicus*, a hyperthermophilic and acidophilic crenarchaeon, revealed an unprecedented level of O-glycosylation on this filament<sup>26</sup>. The adhesive pilus displayed an anomalous amino acid composition, with 37% of the residues in the C-terminal globular domain being threonines and serines, most of which are exposed and likely glycosylated on the surface of the filament. The core of the filament was, as expected, extremely hydrophobic, since the N-terminal domains that constitute the core are transmembrane helices prior to filament assembly. However, the overall percentage of charged residues within the entire pilin protein was only 1.5%. Remarkably, the *S. islandicus* pilus was found to be extremely resilient under various harsh conditions, including extended treatment with different proteases and even boiling in 5 M guanidine hydrochloride. The reason for the robustness of the pilus was attributed to a combination of profound hydrophobicity and extensive surface glycosylation of the filament<sup>26</sup>. Notably, the resilience of the *S. islandicus* pilus precluded the identification of the constituent pilin protein using conventional approaches, such as gel electrophoresis and mass spectrometry. Instead, the identity of the protein was determined using a combination of cryo-EM and bioinformatics, whereby candidate protein sequences were directly threaded through the density maps, which allowed building an atomic model de novo, exemplifying the power of cryo-EM for protein identification<sup>26</sup>.

Whether T4P of other hyperthermophilic archaea display similar sequence features and properties as those described for *S. islandicus* pilus remains unclear. Indeed, the existence of a single atomic structure for an archaeal T4P has limited generalizations about such filaments as well as comparisons with homologs. We present in this paper the structure of two additional archaeal T4P, one from *Pyrobaculum arsenaticum*, a neutrophilic hyperthermophile (optimal growth at 90 °C, pH 7) and the other from *Saccharolobus solfataricus*, an acidophilic hyperthermophile (optimal growth at 75–80 °C, pH 2–3). We use the three structures of archaeal T4P to make comparisons with bacterial T4P and archaeal flagellar filaments. We show that whereas the N-terminal domain is conserved in all bacterial and archaeal T4P-like filaments, the C-terminal domain of bacterial T4P is unrelated to the immunoglobulin (Ig)-like domain shared by archaeal T4P and archaeal flagella. Furthermore, comparison of the pili from the acidophilic and neutrophilic hyperthermophiles suggests that extensive glycosylation is a response to an acidic

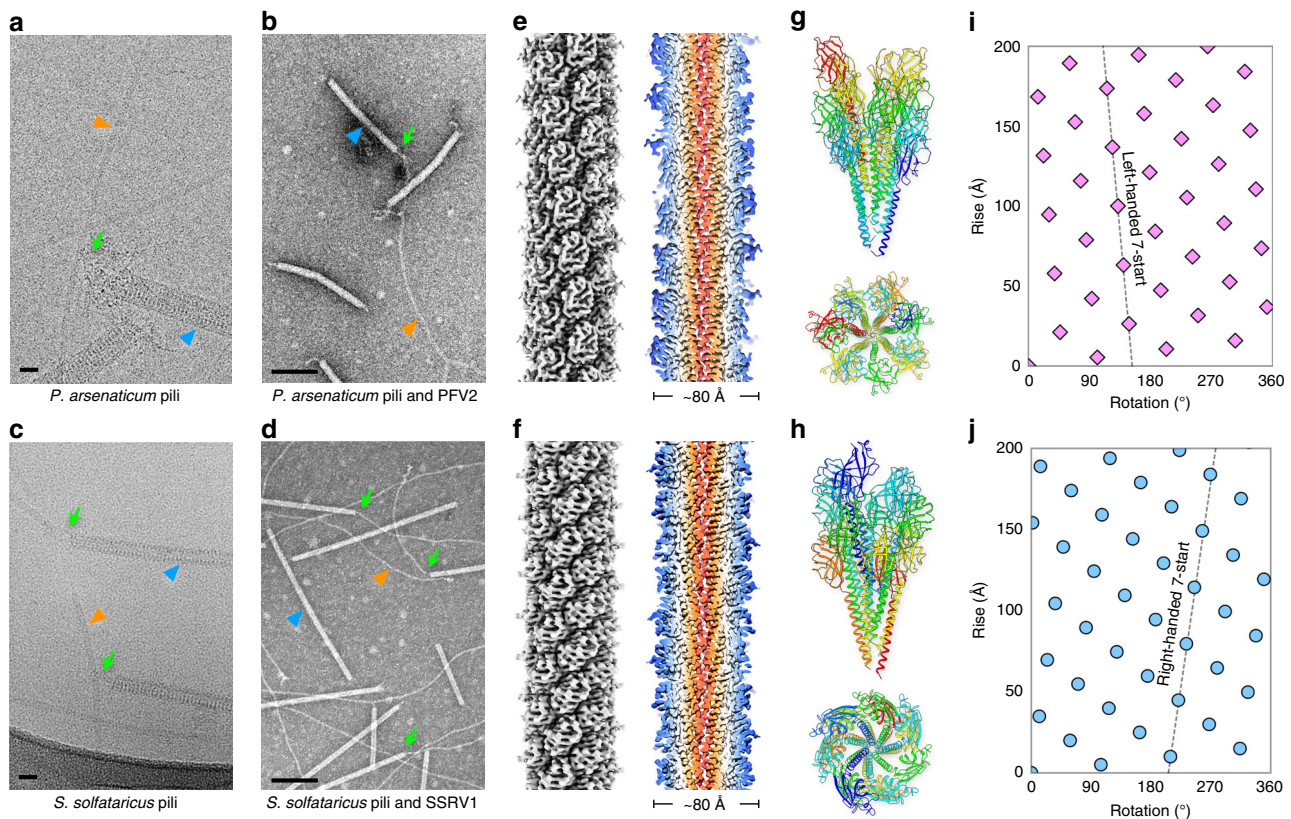
environment, rather than simply high temperature. Notably, the pili of *Pyrobaculum* are much less glycosylated compared to the *Sulfolobus* and *Saccharolobus* pili, and as a consequence, the *Pyrobaculum* filaments do not display the remarkable stability of the *Sulfolobus* filaments in vitro. Finally, we show that *Pyrobaculum* and the *Saccharolobus* T4P are at the forefront of virus–host interaction and are recognized by two different viruses, likely serving as primary receptors during virus infection.

## Results

**Archaeal viruses bind to pili.** We have recently isolated two filamentous viruses, *Saccharolobus solfataricus* rod-shaped virus 1 (SSRV1; family *Rudiviridae*) and *Pyrobaculum* filamentous virus 2 (PFV2; family *Tristromaviridae*), infecting *Saccharolobus solfataricus* POZ149 and *Pyrobaculum arsenaticum* 2GA, respectively<sup>27,28</sup>. After centrifugation in CsCl density gradient, both viruses copurify with and are often bound with their terminal fibers to host-derived pili-like filaments, which were ~80 Å in diameter (Fig. 1a, b). Consequently, the filaments appear to serve as the primary receptors for SSRV1 and PFV2, as has been demonstrated for some other archaeal viruses<sup>16–18</sup>.

**Cryo-EM structures of two archaeal pili.** In the absence of genetic tools for *S. solfataricus* POZ149 and *P. arsenaticum* 2GA, we set out to determine the identity of the SSRV1 and PFV2 host receptors directly by cryo-EM, applying the strategy recently employed for identification of the *S. islandicus* pilin<sup>26</sup>. To this end, mixtures of virions and pili-like filaments were imaged by cryo-EM separately for SSRV1 and PFV2, respectively, and the images of the pili were used for three-dimensional reconstruction. While multiple helical symmetries appeared to be possible based upon analysis of power spectra from these filaments<sup>29</sup>, in each case only one helical symmetry led to a reconstruction with clearly interpretable secondary structure. This was a rise of ~5.3 Å and a rotation of ~101.7° for the *P. arsenaticum* filaments and a rise of ~5.0 Å and a rotation of ~104.6° for the *S. solfataricus* filaments (Table 1). The small difference in helical twist between the two resulted in 7-start long-pitch helices that were left-handed for *P. arsenaticum* (Fig. 1e, i) and right-handed for *S. solfataricus* (Fig. 1f, j). The 7-start helices arise from the fact that there are ~3.5 subunits per turn of the right-handed 1-start helix in both filaments. However, in *P. arsenaticum* pilus, there are slightly more than 3.5 subunits per turn, so there are 7.1 subunits in two turns, while in *S. solfataricus* there are slightly less than 3.5 subunits per turn, yielding 6.9 subunits in two turns. The reconstructions had a resolution of 3.8 Å for *P. arsenaticum* pilus and 3.4 Å for *S. solfataricus* pilus, as judged by a map:map Fourier shell correlation (FSC), a map:model FSC as well as d<sub>99</sub><sup>30</sup> (Supplementary Fig. 1, Table 1).

Hyperthermophilic archaea typically carry multiple pilin genes<sup>13</sup>. Thus, to identify the exact pilin genes encoding receptors of SSRV1 and PFV2, we have sequenced the complete genomes of *S. solfataricus* POZ149 and *P. arsenaticum* 2GA (see Methods), and used a combination of mass spectrometry, bioinformatics, and cryo-EM to build de novo the atomic models (Fig. 2a, b), as was done for the pili from *S. islandicus*<sup>26</sup>. The result in each case was finding a single protein sequence that could be threaded through the maps, which is HC235\_07175 for *P. arsenaticum* (Par\_PilA) and HFC64\_06120 for *S. solfataricus* (Sso\_PilA) (Supplementary Figs. 2, 3). Our structural models indicate that the first 12 residues in Sso\_PilA and 7 residues in Par\_PilA correspond to the signal sequence, which is cleaved from the mature pilin protein. In Sso\_PilA, cleavage occurs between the Gly12 and Leu13 residues (Supplementary Fig. 2d) and the cleavage site matches perfectly with the previously established<sup>11</sup>



**Fig. 1 Cryo-EM of the *P. arsenaticum* pilus and the *S. solfataricus* pilus.** Representative cryo-electron (a) and negative staining (b) micrographs of the *P. arsenaticum* pilus interacting with filamentous virus PFV2. Scale bar, 20 nm in a and 200 nm in b. Orange arrowhead points to the *P. arsenaticum* pilus; blue arrowhead points to the PFV2 virion; green arrow points to the regions of the pilus-virion interaction. Representative cryo-electron (c) and negative staining (d) micrographs of the *S. solfataricus* pilus interacting with filamentous virus SSRV1. Scale bar, 20 nm in c and 200 nm in d. Orange arrowhead points to the *S. solfataricus* pilus; blue arrowhead points to the SSRV1 virion; green arrow points to the regions of the pilus-virion interaction. Cryo-EM reconstruction of the *P. arsenaticum* pilus at 3.8 Å resolution (e) and the *S. solfataricus* pilus at 3.4 Å resolution (f). Thin slices parallel to the helical axis of the pilus are shown, colored by the helical radius. g, h Side view and top view of the *P. arsenaticum* and *S. solfataricus* pilus atomic models, built into the cryo-EM maps shown in e, f. The model is colored by chain. Helical net of the pilus (*P. arsenaticum* pilus in i and the *S. solfataricus* pilus in j) using the convention that the surface is unrolled and we are viewing from the outside.

consensus recognition motif of the signal peptidase PibD ((Lys/Arg)-(Gly/Ala)-↓-(Leu/Ile/Phe)-(Ser/Thr/Ala)). In contrast, in Par\_PilA, cleavage occurs between residues Gly7 and Met8 of the prepilin, and the PibD cleavage site (Arg-Gly↓Met-Thr) slightly deviates from the consensus motif (Supplementary Fig. 2c). BLASTP searches have shown that orthologs of Sso\_PilA are conserved in other members of the order Sulfolobales, including different species of the genera *Saccharolobus*, *Sulfolobus*, *Acidianus* and others. Similarly, Par\_PilA pilins are conserved in *Pyrobaculum* species. Notably, in a number of *Pyrobaculum* genomes, the orthologs of Par\_PilA appear to be misannotated, with the second Met codon (corresponding to Met8 of Par\_PilA prepilin) selected as the start codon; as a result, these proteins appear to lack the signal sequence.

The *S. solfataricus* pilus was very similar to that of *S. islandicus*, albeit at considerably higher resolution (3.4 versus 4.1 Å), which was not surprising given the ~81% sequence identity between the two. In contrast, the sequence identity between the pilins from *P. arsenaticum* and *S. solfataricus* is only ~18% (Fig. 2c). The structural comparison between the two pilins (Fig. 2d) reveals that, even with low sequence identity, both the N-terminal and C-terminal domains are structurally conserved. They both have the archaeal T4P two-domain fold described for the *S. islandicus* LAL14/1 pilus<sup>26</sup>, containing a ~35 amino acid-long N-terminal α-helical domain and C-terminal globular Ig-like β-sandwich

domain. In fact, there is nearly a one-to-one mapping of secondary structural elements between the two (Fig. 2e), with the exception that one β-strand in *P. arsenaticum* (residues 36–46) maps to two shorter β-strands in *S. solfataricus*. However, the relative orientation of these two domains is different by ~40° (Fig. 2d) and, as a result, the helical packing in the filaments is different (Fig. 1g, h). Interestingly, and in contrast to the *S. islandicus* and *S. solfataricus* pilins, *P. arsenaticum* pilin has a 15 amino acid C-terminal extension that wraps around the globular domain. Within this extension, an intramolecular disulfide bond is formed between Cys139 in the extension and Cys80 within the same pilin chain (Fig. 2f, Supplementary Fig. 4), which is likely to contribute to the thermo-stability of this pilus.

**Glycosylation of the pili.** Despite high sequence similarity between the pilins of *S. islandicus* and *S. solfataricus*, SSRV1 does not bind to *S. islandicus* pilus and, accordingly, cannot infect this organism<sup>27</sup>. Given that the surface of the *S. islandicus* pilus is coated with a glycan layer<sup>26</sup>, the latter might play an important, if not key, role in virus binding. Thus, we next analyzed the glycosylation status of the *S. solfataricus* and *P. arsenaticum* pili. It turned out that amino acid distributions and glycosylation represent the most striking differences between the two pilins (Supplementary Table 1). In the *S. solfataricus* pilin, similar to *S.*

**Table 1 Cryo-EM and refinement statistics of the *P. arsenaticum* and the *S. solfataricus* pilus.**

| Parameter                                           | <i>P. arsenaticum</i> pilus | <i>S. solfataricus</i> pilus |
|-----------------------------------------------------|-----------------------------|------------------------------|
| Data collection and processing                      |                             |                              |
| Voltage (kV)                                        | 300                         | 300                          |
| Electron exposure (e <sup>-</sup> Å <sup>-2</sup> ) | 50                          | 58                           |
| Pixel size (Å)                                      | 1.08                        | 1.08                         |
| Final particle images (n)                           | 210,331                     | 579,713                      |
| Helical symmetry                                    |                             |                              |
| Point group                                         | C1                          | C1                           |
| Helical rise (Å)                                    | 5.26                        | 4.97                         |
| Helical twist (°)                                   | 101.69                      | 104.57                       |
| Map resolution (Å)                                  |                             |                              |
| Map:map FSC (0.143)                                 | 3.8                         | 3.4                          |
| Model:map FSC (0.38)                                | 3.8                         | 3.4                          |
| d <sub>99</sub>                                     | 3.9                         | 3.6                          |
| Refinement and model validation                     |                             |                              |
| Map-sharpening B-factor (Å <sup>2</sup> )           | -140                        | -104                         |
| Bond lengths RMSD (Å)                               | 0.005                       | 0.006                        |
| Bond angles RMSD (°)                                | 1.004                       | 0.986                        |
| Model:map RSCC                                      | 0.87                        | 0.85                         |
| Clashscore                                          | 8.82                        | 7.28                         |
| Poor rotamers (%)                                   | 0.5                         | 0                            |
| Ramachandran favored (%)                            | 95.63                       | 95.73                        |
| Ramachandran outlier (%)                            | 0                           | 0                            |
| MolProbity score                                    | 1.78                        | 1.70                         |
| Deposition ID                                       |                             |                              |
| PDB (model)                                         | 6W8U                        | 6W8X                         |
| EMDB (map)                                          | EMD-21578                   | EMD-21579                    |

*islandicus*, only about 1.5% of the residues are charged. We showed<sup>26</sup> that extensive glycosylation makes these extremely hydrophobic filaments soluble. In contrast, ~10% of the residues in *P. arsenaticum* pilin are charged. In *S. solfataricus* and *S. islandicus*, ~37% of the residues in the globular domain are either serine or threonine, potential sites for O-linked glycosylation. In *P. arsenaticum* pilin as well as archaeal flagella and adhesion filaments<sup>22–26</sup>, only ~20% of the residues are serine or threonine (Fig. 3a). No extra density corresponding to O-linked glycosylation was reported, except for one serine in the flagellar filament of *M. hungatei*<sup>22</sup>. In *S. solfataricus*, we see clear extra density consistent with glycosylation on seven residues, all serines or threonines (Fig. 3b). In contrast, in *P. arsenaticum* only a single density consistent with glycosylation is seen, on Asn55 (Fig. 3b), which would be N-linked glycosylation.

We expect that most glycosylation is not seen in the high-resolution maps due to the disorder of the sugars, and the fact that the same residue may not be glycosylated in every subunit. However, we can directly visualize at lower resolution additional mass consistent with glycosylation (Fig. 3c) by subtracting the predicted density due to the atomic model from the actual maps, after both are filtered to 7 Å. We are fortunate to have a control for this procedure which involves the rod-shaped virus SSRV1 present in the images for the *S. solfataricus* pili. There was no apparent glycosylation of this virus, and the small extra density not explained by the atomic model involves a few disordered C-terminal residues of the capsid protein that were not part of the atomic model. It can be seen (Fig. 3c) that, as expected, the *S. solfataricus* filament has a significantly greater amount of peripheral density due to glycosylation than the *P. arsenaticum* filament. A prediction of the lesser glycosylation for the *P. arsenaticum* pilus is that it will not display the remarkable stability of the *S. islandicus* pilus in vitro<sup>26</sup>. Consistent with this prediction, boiling the *S. islandicus* pili in 5 M GuHCl failed to disassemble the filaments<sup>26</sup>, but under the same conditions the *P. arsenaticum* pili come apart (Supplementary Fig. 5). Collectively, these results suggest that SSRV1 and PFV2 recognize their corresponding receptors through different interactions. Whereas binding of SSRV1 to *S. solfataricus* pili is likely to be mediated

through carbohydrate moieties, direct interaction with the pilin subunits might play an important role in the case of PFV2.

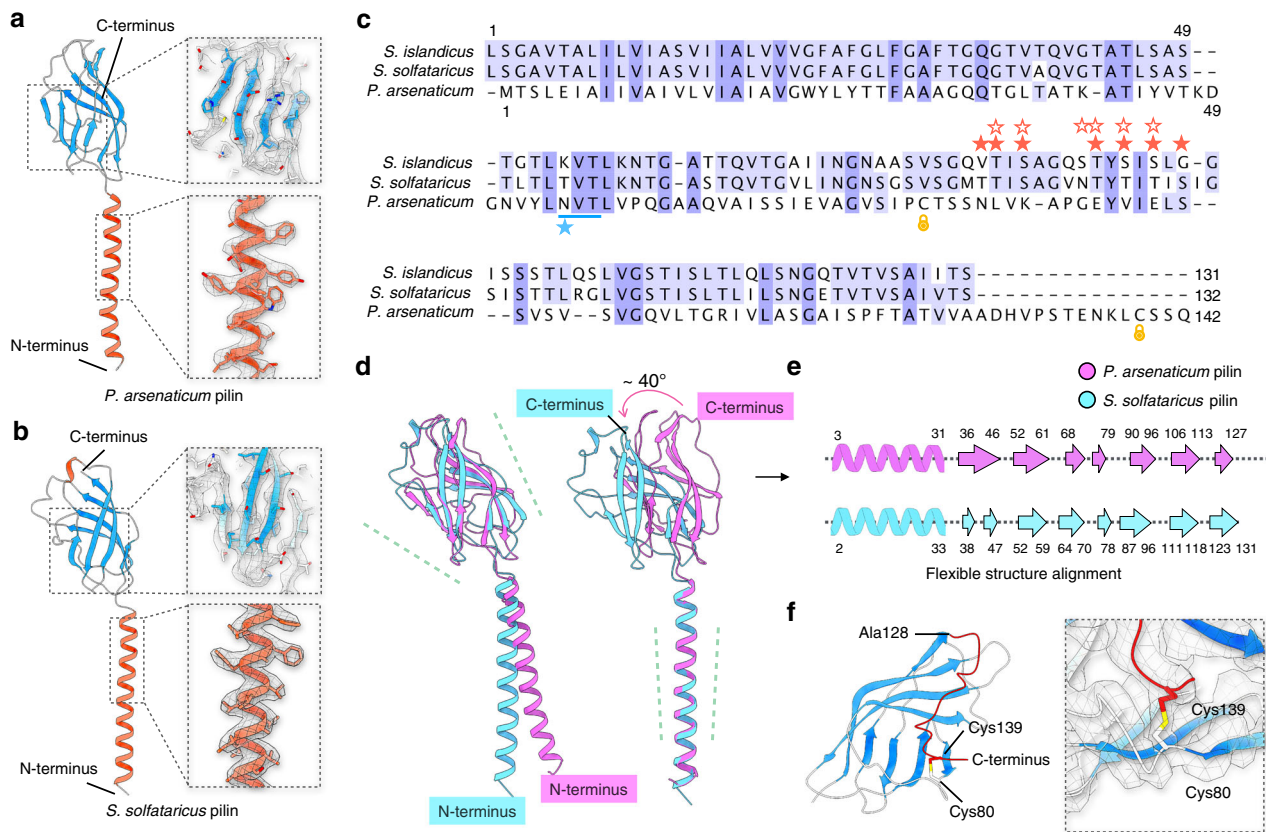
**Evolutionary relationships.** Expanding the number of archaeal T4P atomic structures from one to three allows for more reliable comparisons to now be made between these structures and both bacterial T4P and archaeal flagellar filaments. The most striking difference is that in both bacterial T4P and archaeal flagellins, the N-terminal  $\alpha$ -helix is ~50 residues long, while in the archaeal T4P it is only ~35 residues long (Fig. 4a). By this criterion, the Iho670 adhesion filaments, previously described as “flagellar-like”<sup>25</sup>, appear to be a form of archaeal T4P. Despite the differences in the N-terminal helix length between archaeal flagellins and T4P, the helical symmetries of the filaments are quite similar (Fig. 4b). In contrast, although bacterial T4P can have a similar twist to the archaeal filaments (~100°), they have a distinctly different axial rise of ~10 Å as opposed to the ~5 Å in the archaeal filaments. In addition, all known bacterial T4P show a partial melting of the N-terminal helix, not seen in any of the archaeal filaments<sup>31</sup>.

The distribution of helical symmetries among the archaeal filaments and bacterial T4P also suggests a conservation of how their long N-terminal  $\alpha$ -helical tails pack together (Fig. 4b). Therefore, five adjacent copies of those  $\alpha$ -helical tails were cut out from the filament model and compared with each other (Fig. 4c, d). Interestingly, Iho670 adhesion filament packs very similarly to *S. solfataricus* and *S. islandicus* pili, with an RMSD < 2 Å, while *P. arsenaticum* packs slightly differently from those three filaments with RMSDs of 3–4 Å (Fig. 4d). Notably, the crucial N-terminal subunit-subunit interactions in these four filaments are almost all hydrophobic (Fig. 4a, c), except for in *P. arsenaticum* there is a glutamic acid, which may contribute to the slightly different packing compared to Iho670 and other archaeal T4P. The archaeal flagellar filaments, judging by the sequence alignment and the RMSD matrix (Fig. 4a, d), have a very similar packing with Iho670 and the other archaeal T4Ps in their first 35 amino acids. However, their helices extend further with more hydrophilic residues interacting both with each other and with the globular domains, presumably providing greater mechanical stability in their roles as rotating propellers. Bacterial T4Ps, in contrast, have more hydrophilic residues in their N-terminal helices, including the conserved glutamic acid-forming key hydrogen bonds with N-terminal residues. Because of the melted N-terminal region seen thus far in every bacterial T4P, they have a completely different packing compared to archaeal filaments (Fig. 4c, d).

A DALI<sup>32</sup> matrix analysis of the C-terminal globular domains (Fig. 5a) shows that archaeal flagellins, T4P and the Iho670 adhesion filament share structural homology for this domain containing an Ig-like  $\beta$ -sandwich. However, this domain in archaea shows no structural similarity to the bacterial T4P, where this globular domain is mainly a  $\beta$ -sheet containing 3–5  $\beta$ -strands, or almost entirely absent in at least one pilin<sup>33</sup>. A similar matrix was seen when the C-terminal globular domains were analyzed by TM-score<sup>34</sup> (Supplementary Fig. 6). TM-score is a metric for measuring the similarity of protein structures, and has a value in the range (0,1). It is a better metric in this instance, because it is independent of protein length, unlike RMSD analysis. Therefore, archaeal T4P and flagellins cannot be readily distinguished by the fold of the C-terminal domain, but can be easily distinguished simply by the length of the N-terminal helix.

## Discussion

Recent advances in cryo-EM have revolutionized the life sciences<sup>35–37</sup>. Besides yielding near-atomic resolution structures of complex macromolecular assemblies, this technique becomes



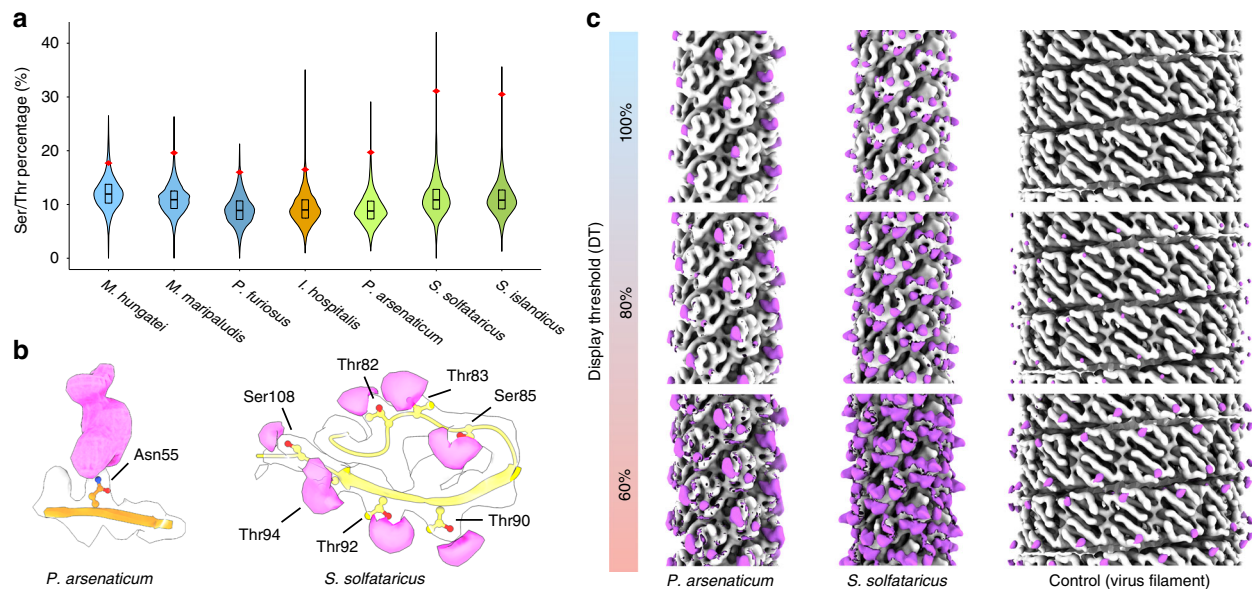
**Fig. 2 Two-domain architecture of archaeal type IV pilus (T4P).** Protein C $\alpha$  trace of a single *P. arsenaticum* pilin (**a**) and *S. solfataricus* pilin (**b**). Pilin structures are shown in cartoon, colored by secondary structure. The map quality of the N-terminal long helix and C-terminal globular domain are shown on the right. **c** Structure-based sequence alignments of three archaeal type IV pilins with known structures, *P. arsenaticum* pilin, *S. solfataricus* pilin and previously determined *S. islandicus* pilin. Sequences were aligned using Tcoffee<sup>67</sup> and visualized with Jalview<sup>68</sup>. The two blue shades correspond to the amount of sequence identity (dark blue when all three are identical, light blue when two are identical). The observed glycosylated residues are indicated with stars: solid red stars for O-linked sugars in *S. solfataricus*, open red stars for O-linked sugars in *S. islandicus*, and a blue star for N-linked sugar in *P. arsenaticum*. The blue line above the star indicates the NxS/T motif associated with N-linked glycosylation in *P. arsenaticum*. Two cysteine residues that form a disulfide bond in *P. arsenaticum* are indicated with yellow lock symbols. Note that amino acid positions are counted starting with the first amino acid of the mature protein (i.e., following cleavage of the signal peptide). **d** Superposition of a single *P. arsenaticum* pilin (purple) and *S. solfataricus* pilin (cyan), aligned by either the globular domain (left) or the N-terminal helix (right). **e** Flexible secondary structure alignment reveals a simple mapping of the local secondary structure elements between the two pilins. The N-terminal domain is  $\alpha$ -helical, and each  $\beta$ -strand in the C-terminal domain is shown as a closed shape. **f** The C-terminal loop of the *P. arsenaticum* pilin and the disulfide bond.

increasingly valuable for de novo identification and characterization of protein components constituting these assemblies, whereby atomic models are built directly into the map densities. This approach has been particularly revealing in the study of complex virions<sup>38–41</sup> and cellular appendages<sup>26</sup>. Here we have used cryo-EM to identify the host receptors recognized by two hyperthermophilic archaeal viruses, SSRV1 and PFV2. Studies on virus-host interactions remain highly challenging in archaea, due to the lack of genetic tools for most organisms, as is the case for the hosts of SSRV1 and PFV2. We have observed that both viruses bind to pili-like filaments and could determine the identity of the respective receptors directly by cryo-EM. We do not know how this would be possible using other approaches. For instance, mass-spectrometry analysis of a crude sample typically yields too many potential candidate proteins for unequivocal identification, whereas some proteins are not amenable to standard tryptic digestion, as has been shown recently for the pili of *S. islandicus*<sup>26</sup>.

Pili appear to be a common receptor for hyperthermophilic archaeal viruses and have been previously shown to be recognized by rudiviruses SIRV2 and SIRV8<sup>16,17</sup>, and turdivirus STIV<sup>18</sup>. Here we have shown that members of yet another family of archaeal

viruses, the *Tristromaviridae*, also bind to the host cells via T4P. Rudiviruses recognize the adhesive T4P<sup>26</sup>, whereas the identity of the pilus receptor of STIV on the surface of *S. solfataricus* has not been determined<sup>18</sup>. We do not exclude the possibility that the *S. solfataricus* T4P identified herein as the receptor for SSRV1 is an ortholog of the pilus recognized by STIV. Notably, direct adsorption experiments with SSRV1 to different members of the genera *Sulfolobus* and *Saccharolobus* showed that the virus could adsorb only to *S. solfataricus* POZ149 and, less efficiently, to *S. solfataricus* P1, but not to *S. islandicus* LAL14/1 or *S. acidocaldarius* DSM 639<sup>27</sup>, suggesting that efficient binding is highly specific. Intriguingly, in addition to the divergence of the pilin proteins themselves, there might be species-specific differences in the pilin glycosylation patterns, which could play equally important role in determining the host range of archaeal viruses. Thus, structural characterization of the glycans decorating different archaeal surface structures is likely to shed additional light on virus-host interactions in archaea.

The reasons underlying the selection of T4P as common virus receptors remain unclear. One possible explanation is that T4P are important for survival and functioning of hyperthermophilic archaea in natural habitats and hence the loss of these structures



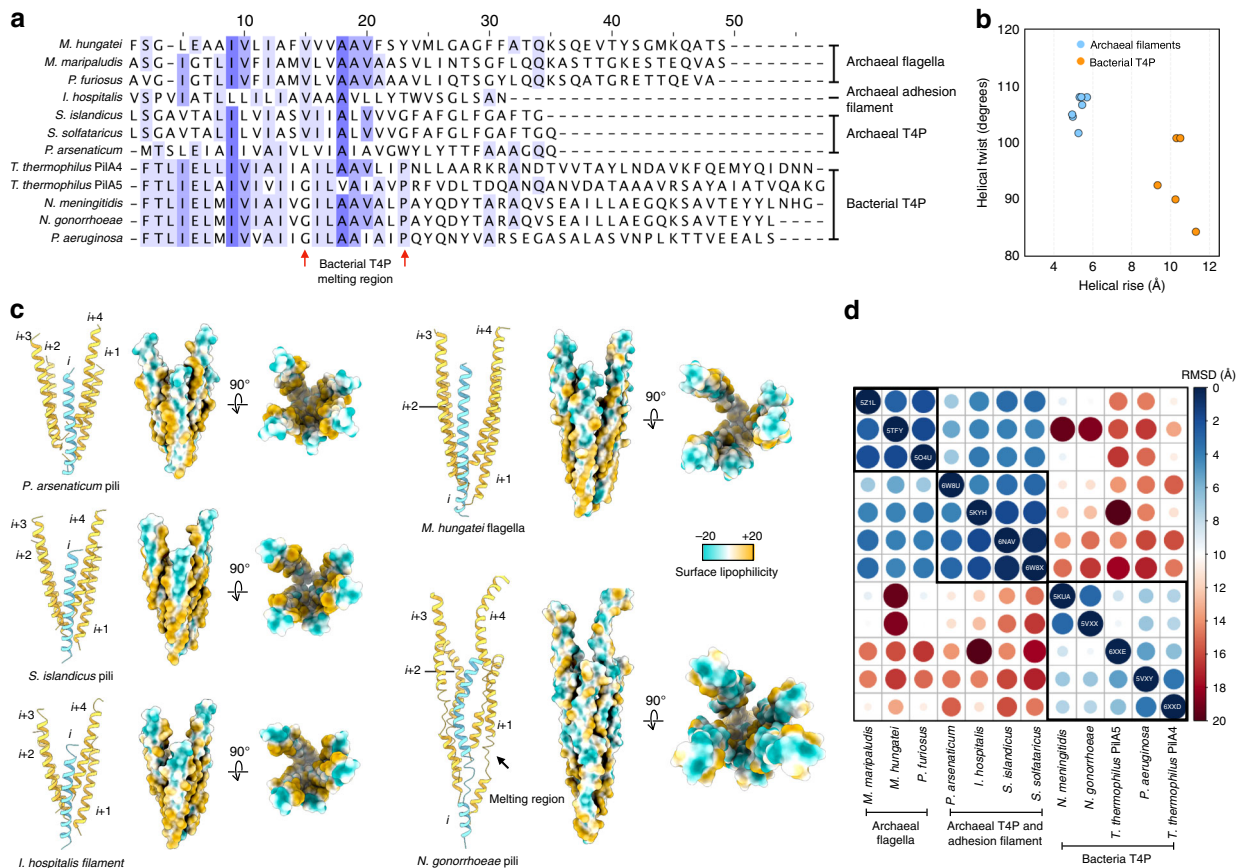
**Fig. 3 Glycosylation of archaeal type IV pilus and other archaeal cell appendages.** **a** Violin plot showing the distributions of the serine and threonine percentages in all proteins of *M. hungatei*, *M. maripaludis*, *P. furiosus*, *I. hospitalis*, *P. arsenaticum*, *S. solfataricus* and *S. islandicus*. The red diamond indicates the percentage of serine+threonine in the pilin or flagellin from this species discussed in this paper. The black box indicates the interquartile range, and the central line in the box indicates the median of the data. The species for the archaeal flagellar filaments are colored in blue-ish colors, adhesion filaments in orange, and archaeal T4P in green-ish colors. **b** Density (magenta) due to post-translational modifications on *P. arsenaticum* pilin (left) and *S. solfataricus* pilin (right). The protein C $\alpha$  backbone is colored in orange (*P. arsenaticum*) or yellow (*S. solfataricus*). The extra density is due to N-linked (Asn) or O-linked (Thr, Ser) sugars, and the residues modified are labeled. **c** Surface glycosylation of the *P. arsenaticum* pilus (left), *S. solfataricus* pilus (middle) and a control rod-like virus without apparent glycosylation. The control virus and the *S. solfataricus* pilus were purified and imaged together. All three volumes were filtered to 7 Å. The density accounted for by atomic models is colored in gray, and the extra density is colored in magenta. Three different display thresholds are shown. A small additional density appears on the surface of the virus, and this is due to several C-terminal residues that were not part of the atomic model due to disorder, and not due to glycosylation.

would bear a considerable fitness cost. Consistently, T4P are highly widespread in archaea and generally well conserved in members of the same genus<sup>13</sup>. Notably, it has been shown that besides its host *S. solfataricus* POZ149, SSRV1 can deliver its genome, with low efficiency, into a broad range of cells from genera *Saccharolobus*/*Sulfolobus* and *Acidianus*, even though genome delivery did not lead to virus proliferation<sup>27</sup>. An alternative explanation could be that the inherent properties of T4P might assist, directly or otherwise, in the genome delivery into the cell interior. For instance, T4P might be landmarks for the openings in the S-layer, an external crystalline protein layer surrounding archaeal cells<sup>42,43</sup>, which has to be penetrated during the delivery of the viral genome. Furthermore, it has been previously suggested that SIRV2 initially binds to the tips of T4P and progressively moves along the pilus to the cell body, where genome delivery takes place<sup>16</sup>. However, the mechanisms which would promote the virus movement and genome internalization remain to be understood. Additionally, if archaeal T4P retract, as they do in bacteria<sup>1</sup>, they would be ideal candidates as receptors for viruses as they could readily lead to the internalization of the viral genome. However, no PilT-like retraction ATPases have been found in archaea, and no direct observations of T4P retraction in archaea have been made thus far<sup>13</sup>. But the Tad pili in *Caulobacter*, which also lack a PilT-like retraction ATPase, have now been shown to retract<sup>44</sup>. We therefore leave this possibility of pilus retraction open.

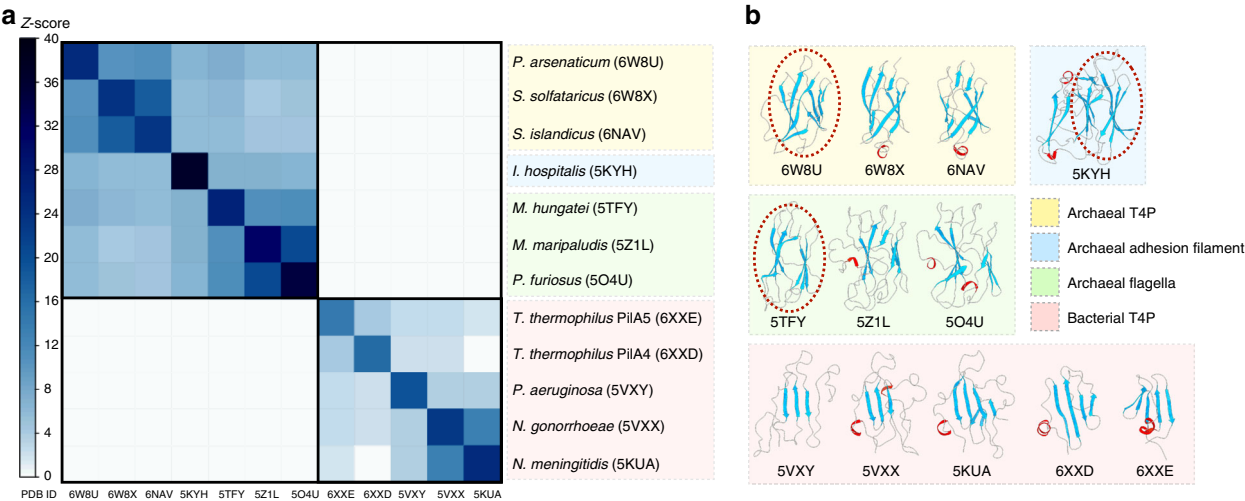
Being exposed on the cell surface, T4P of hyperthermophilic archaea represent an excellent model for studying the remarkable stability of macromolecular complexes under extreme physico-chemical conditions. Previous study of the *S. islandicus* T4P has revealed that extreme hydrophobicity coupled with extensive glycosylation renders the pilus resistant to various harsh

treatments, including boiling in SDS and GuHCl or prolonged treatment with proteases<sup>26</sup>. Comparison of T4P from different hyperthermophiles allows us to dissect further the factors underlying adaptations to different environmental conditions. First, although T4P of a neutrophilic *P. arsenaticum* is as hydrophobic as those of acidophilic *S. islandicus* and *S. solfataricus* (Fig. 2c), it does not possess the extensive sugar coating. This observation suggests that glycosylation plays a more important role for survival in acidic environments. Recently, another adaptation to low pH habitats has been uncovered in archaea-specific glycerol dibiphytanyl glycerol tetraethers (GDGTs) lipids. Headgroups of GDGT in acidophilic hyperthermophiles are covalently linked to calditol, a unique five-membered ring, which is synthesized from a hexose sugar moiety<sup>45</sup>. Decoration of pili and lipids with covalently linked glycans and calditol, respectively, appear to be conceptually similar adaptations to acidic environments. Instead, T4P of *P. arsenaticum* is stabilized by an intramolecular disulfide bond (Fig. 2f), a protein adaptation to high temperatures previously reported in bacteria, archaea and their viruses<sup>46–48</sup>. Disulfides in mesophilic bacterial T4P have also been previously described<sup>49</sup>. Notably, qualitative observations suggest that *P. arsenaticum* T4P do not possess the remarkable resilience observed for *Sulfolobus* T4P (Supplementary Fig. 5), suggesting that it may be glycosylation rather than hydrophobicity which renders the pilus nearly indestructible.

While it was previously suggested that the Iho670 adhesion filaments from the non-motile archaeon *I. hospitalis* were “flagellar-like”<sup>25</sup>, we have now shown that all archaeal T4P (or at least, the three structures that now exist) are truly flagellar-like in comparison with the atomic structures for archaeal flagellar filaments. Further, the Iho670 filaments are much closer to



**Fig. 4 Conservation and divergence of the N-terminal helical bundles.** **a** Sequence alignments of the N-terminal long helix in three archaeal flagellins, one archaeal “adhesion” protein, three archaeal T4P and three bacterial T4P. Area where the bacterial T4P N-terminal helix melts is indicated by two red arrows. **b** Helical symmetries of the filaments mentioned in **a**, with archaeal filaments light blue and bacterial filaments orange. **c** Five adjacent N-terminal helices of *P. arsenaticum* pilins, *S. solfataricus* pilins, *I. hospitalis* proteins, *M. hungatei* flagellins and *N. gonorrhoeae* pilins are shown. The Cα trace is shown in cartoon (left), with one helix colored cyan and the other four colored yellow. The bundles are also shown in an atomic surface view and colored by lipophilicity (right). **d** All-against-all comparison of the five helix bundles of the structures shown in **a**. The matrix is based on the pairwise RMSD comparisons calculated from PyMOL.



**Fig. 5 Conservation and divergence of the C-terminal globular domains.** **a** All-against-all comparison of the bacterial T4P and other archaeal structures with N-terminal T4P domains. The N-terminal helices have been removed and only the globular domains were analyzed. The matrix is based on the pairwise Z-score comparisons calculated using the DALI server<sup>32</sup>. The color scale indicates the corresponding Z-scores. Based on their functions and origins, the proteins are partitioned into four groups, which are indicated with different colors on the right. One representative globular domain within each archaeal group was highlighted by a red dashed circle, to show the corresponding homology region. **b** Ribbon diagrams of the globular domains used in **a**.

archaeal T4P than to flagellar filaments, particularly in terms of the N-terminal helix that fails to extend into the C-terminal globular domain. Thus, we think it likely that there may be many other archaeal proteins that are not identified at this point as T4P. Given that the C-terminal domains of T4P-like filaments in bacteria and archaea are unrelated and some bacteria lack this domain altogether<sup>33</sup>, it appears likely that the N-terminal domain consisting of the signal sequence and the hydrophobic  $\alpha$ -helix, which forms the core of the filament, has evolved first, possibly at the stage of the last universal cellular ancestor (LUCA). In contrast, the C-terminal globular domains have emerged later on, independently in Bacteria and Archaea. Consistent with this, a novel T4P fold has now been found in bacteria, where the C-terminal domain is all  $\alpha$ -helical<sup>50</sup>. The archaeal C-terminal domain is clearly related to immunoglobulin-like domains which are widespread in all walks of life<sup>51</sup>. Thus, the contemporary archaeal T4P-like filaments are a result of a fusion of two pre-existing structural domains, a common mechanism of protein evolution and diversification<sup>52,53</sup>. Importantly, a close structural similarity between archaeal pili and flagellar filaments (Fig. 5) strongly suggests that functional diversification of the two types of archaeal filaments took place following the split of bacterial and archaeal lineages from their common ancestor.

## Methods

**Strain cultivation and filament purification.** Exponentially growing cultures of *Saccharolobus solfataricus* POZ149<sup>27</sup> and *Pyrobaculum arsenaticum* 2GA<sup>54</sup> cells were infected with fresh preparations of *Saccharolobus solfataricus* rod-shaped virus 1 (SSRV1) and *Pyrobaculum filamentous virus* 2 (PFV2), respectively. The infected culture of *S. solfataricus* POZ149 was incubated at 75 °C under agitation for 2 days, whereas *P. arsenaticum* 2GA culture was grown during 4 days at 90 °C without shaking. After the removal of cells (7438  $\times$  g, 20 min, Sorvall 1500 rotor), viruses along with pili detached from the cells were collected and concentrated by ultracentrifugation (273,620  $\times$  g, 2 h, 10 °C, Beckman 126 SW41 rotor). For cryo-EM analysis, the concentrated particles were resuspended in buffer A<sup>55</sup>: 20 mM KH<sub>2</sub>PO<sub>4</sub>, 250 mM NaCl, 2.14 mM MgCl<sub>2</sub>, 0.43 mM Ca(NO<sub>3</sub>)<sub>2</sub> and <0.001% trace elements of Sulfolobales medium, pH 6. Virus particles and pili also copurified when the pelleted virion-pili mixtures were subjected to centrifugation in a CsCl buoyant density gradient (0.45 g ml<sup>-1</sup>) with a Beckman SW41 rotor at 260,110  $\times$  g for 20 h at 10 °C.

**DNA extraction, sequencing and genome analysis.** DNA was extracted from exponentially growing cultures of *S. solfataricus* POZ149 and *P. arsenaticum* 2GA using the Wizard Genomic DNA Purification Kit (Promega). The genome sequence of *S. solfataricus* POZ149 was obtained using PacBio single-molecule real-time (SMRT) technology, while libraries with 250 bp insert size were sequenced on Illumina platform for *P. arsenaticum* 2GA (Novogen, China). Raw sequence reads were processed and assembled with FALCON<sup>56</sup> and SOAPdenovo version 2.04<sup>57</sup>, respectively. Gene prediction was performed by GeneMark.hmm version 3.25<sup>58</sup>. The genome sequence of *S. solfataricus* POZ149 was deposited in GenBank under the accession number CP050869. The Whole Genome Shotgun project for *P. arsenaticum* 2GA has been deposited at DDBJ/ENA/GenBank under the accession JAAVJF000000000. The version described in this paper is version JAAVJF010000000.

**Negative-stain electron microscopy.** For TEM, 5  $\mu$ l of the samples were applied to carbon-coated copper grids and negatively stained with 2% uranyl acetate. The stained grids were imaged with the transmission electron microscope FEI Tecnai Biotwin (Institut Pasteur, France).

**Cryo-electron microscopy and image analysis.** The pilus sample (~4.0  $\mu$ l) was applied to discharged lacey carbon grids and plunge frozen using a Vitrobot Mark IV (FEI). Frozen grids were imaged in a Titan Krios at 300 keV and recorded with a K3 camera at 1.08 Å/pixel. The *P. arsenaticum* pilus data were collected in the UVA core facility, and the *S. solfataricus* pilus was collected in National Cryo-EM Facility (NCEF) of the National Cancer Institute (NCI). Micrographs were collected using a defocus range of 1–2  $\mu$ m, with a total exposure dose of ~50 electrons/Å<sup>2</sup> distributed into ~25 fractions. To get a preliminary helical reconstruction volume in SPIDER<sup>59</sup>, all of the micrographs were first motion corrected and dose weighted by MotionCorr v2, and then CTF-multiplicated by the theoretical CTF. Filament images amounting to ~20 electrons/Å<sup>2</sup> were extracted by EMAN2<sup>60</sup>. A small subset containing ~50,000 overlapping 384-pixel-long segments was used to search for the correct helical symmetry. For *P. arsenaticum* we explored a multiplicity of

symmetries that were each consistent with the averaged power spectrum. The number of possible symmetries was six (Supplementary Table 2), and each of these was tested in SPIDER<sup>59</sup> using IHRSR<sup>61</sup>, until recognizable protein features were seen, such as side-chains. A ~4 Å reconstruction was generated from this small subset with the correct symmetry (entry #1, Supplementary Table 2), and this volume was subsequently filtered to 7 Å as the starting reference used in RELION<sup>62</sup>. For *S. solfataricus*, the high degree of sequence identity with the *S. islandicus* LAL14/1 pilus suggested that a helical symmetry close to that previously determined<sup>26</sup> for *S. islandicus* might be the correct one, and this turned out to be the case. After using the full dataset in RELION (210,331 segments for the *P. arsenaticum* pilus, 579,713 segments for the *S. solfataricus* pilus, Table 1), doing CTF-refinement and Bayesian polishing, the final volumes were estimated to have a resolution of 3.8 Å for the *P. arsenaticum* pilus and 3.4 Å for the *S. solfataricus* pilus, based on the map:map FSC, model:map FSC and  $d_{99}$ <sup>30</sup>. The final volumes were then sharpened with a negative B-factor automatically estimated in RELION (Table 1).

**Model building.** The density corresponding to a single *S. solfataricus* pilin was segmented from the experimental cryo-EM density map using Chimera<sup>63</sup>. Then a *S. solfataricus* pilin model was generated by homology modeling using the *S. islandicus* LAL14/1 pilin as the reference, and then docked into the segmented map. Then this model was adjusted manually in Coot and real-space refined in PHENIX<sup>64</sup>. Finally, the refined single pilin model was used to generate a filamentous model using the determined helical symmetry, and this filament model was refined against the full cryo-EM map using in PHENIX. MolProbity<sup>65</sup> was used to evaluate the quality of the filament model. The refinement statistics are shown in Table 1.

The density corresponding to a single *P. arsenaticum* pilin was also segmented in the same manner. Due to the uncertainty in the pilin sequence, seven potential pilin candidates were selected based on secondary structure predictions (WP\_011899639, WP\_011900226, WP\_011899654, WP\_011900935, WP\_011900944, WP\_011900938, and WP\_011900942). Each of the sequences was used to do de novo model building against the segmented map using Rosetta CM<sup>66</sup>. As a result, only HC235\_07175 could be built into the cryo-EM map and was also detected by MS/MS analysis. To further validate the sequence of both pilin, the whole genomes of both strains were sequenced and deposited. The model building for the assembled *P. arsenaticum* pilus was the same as described above.

**Mass spectrometry.** The solution samples containing the pili were reduced with 10 mM DTT in 0.1 M ammonium bicarbonate then alkylated with 50 mM iodoacetamide in 0.1 M ammonium bicarbonate (both room temperature for 0.5 h). The sample was then digested overnight at 37 °C with 0.1  $\mu$ g (Chymotrypsin, Asp-N and Trypsin) in 50 mM ammonium bicarbonate. The sample was acidified with acetic acid to stop digestion and then spun down. The solution was evaporated to 15  $\mu$ l for MS analysis.

The LC-MS system consisted of a Thermo Electron Q Exactive HF-X mass spectrometer system with an Easy Spray ion source connected to a Thermo 75  $\mu$ m  $\times$  15 cm C18 Easy Spray column. 7  $\mu$ l of the extract was injected and the peptides eluted from the column by an acetonitrile/0.1 M formic acid gradient at a flow rate of 0.3  $\mu$ l/min over 1.0 hours. The nanospray ion source was operated at 1.9 kV. The digest was analyzed using the rapid switching capability of the instrument acquiring a full scan mass spectrum to determine peptide molecular weights followed by product ion spectra (10 HCD) to determine amino acid sequence in sequential scans. This mode of analysis produces approximately 25,000 MS/MS spectra of ions ranging in abundance over several orders of magnitude. Not all MS/MS spectra are derived from peptides. The data were analyzed by database searching using the Sequest search algorithm against Uniprot *Sulfolobus solfataricus* or Uniprot *Pyrobaculum arsenaticum*.

## Data availability

The genome sequence of *S. solfataricus* POZ149 was deposited in GenBank under the accession number CP050869. The Whole Genome Shotgun project for *P. arsenaticum* 2GA has been deposited at DDBJ/ENA/GenBank under the accession number JAAVJF000000000. The version described in this paper is JAAVJF010000000. The atomic model for the *P. arsenaticum* pilus was deposited at the Protein Data Bank with accession code 6W8U, and the corresponding map was deposited at the Electron Microscopy Data Bank with code EMD-21578. The atomic model for the *S. solfataricus* pilus was deposited at the Protein Data Bank with accession code 6W8X, and the corresponding map was deposited at the Electron Microscopy Data Bank with code EMD-21579. Other data are available from the corresponding authors upon reasonable request.

Received: 9 April 2020; Accepted: 22 June 2020;

Published online: 09 July 2020

## References

- Craig, L., Forest, K. T. & Maier, B. Type IV pili: dynamics, biophysics and functional consequences. *Nat. Rev. Microbiol.* **17**, 429–440 (2019).
- Giltner, C. L., Nguyen, Y. & Burrows, L. L. Type IV pilin proteins: versatile molecular modules. *Microbiol. Mol. Biol. Rev.: MMBR* **76**, 740–772 (2012).
- Lu, S. et al. Nanoscale pulling of type IV pili reveals their flexibility and adhesion to surfaces over extended lengths of the pili. *Biophys. J.* **108**, 2865–2875 (2015).
- Craig, L., Pique, M. E. & Tainer, J. A. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**, 363–378 (2004).
- Daum, B. & Gold, V. Twitch or swim: towards the understanding of prokaryotic motion based on the type IV pilus blueprint. *Biol. Chem.* **399**, 799–808 (2018).
- Chang, C. Y. Surface sensing for biofilm formation in *Pseudomonas aeruginosa*. *Front Microbiol.* **8**, 2671 (2017).
- Piepenbrink, K. H. DNA uptake by type IV filaments. *Front Mol. Biosci.* **6**, 1 (2019).
- Bayley, D. P. & Jarrell, K. F. Further evidence to suggest that archaeal flagella are related to bacterial type IV pili. *J. Mol. Evol.* **46**, 370–373 (1998).
- Faguy, D. M., Jarrell, K. F., Kuzio, J. & Kalmokoff, M. L. Molecular analysis of archaeal flagellins: similarity to the type IV pilin-transport superfamily widespread in bacteria. *Can. J. Microbiol.* **40**, 67–71 (1994).
- Albers, S. V. & Pohlschroder, M. Diversity of archaeal type IV pilin-like structures. *Extremophiles* **13**, 403–410 (2009).
- Albers, S. V., Szabo, Z. & Driessen, A. J. Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J. Bacteriol.* **185**, 3918–3925 (2003).
- Albers, S. V. & Jarrell, K. F. The archaeellum: an update on the unique archaeal motility structure. *Trends Microbiol.* **26**, 351–362 (2018).
- Makarova, K. S., Koonin, E. V. & Albers, S. V. Diversity and evolution of type IV pili systems in archaea. *Front Microbiol.* **7**, 667 (2016).
- Pohlschroder, M. & Esquivel, R. N. Archaeal type IV pili and their involvement in biofilm formation. *Front Microbiol.* **6**, 190 (2015).
- Chaudhury, P., Quax, T. E. F. & Albers, S. V. Versatile cell surface structures of archaea. *Mol. Microbiol.* **107**, 298–311 (2018).
- Quemin, E. R. et al. First insights into the entry process of hyperthermophilic archaeal viruses. *J. Virol.* **87**, 13379–13385 (2013).
- Rowland, E. F., Bautista, M. A., Zhang, C. & Whitaker, R. J. Surface resistance to SSVs and SIRVs in pilin deletions of *Sulfolobus islandicus*. *Mol. Microbiol.* <https://doi.org/10.1111/mmi.14435> (2019).
- Hartman, R. et al. The molecular mechanism of cellular attachment for an archaeal virus. *Structure* **27**, 1634–1646 (2019).
- Wang, F. et al. Cryoelectron microscopy reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* type IV pili at sub-nanometer resolution. *Structure* **25**, 1423–1435 (2017).
- Kolappan, S. et al. Structure of the *Neisseria meningitidis* Type IV pilus. *Nat. Commun.* **7**, 13015 (2016).
- Neuhaus, A. et al. Cryo-electron microscopy reveals two distinct type IV pili assembled by the same bacterium. *Nat. Commun.* **11**, 2231 (2020).
- Poweit, N. et al. CryoEM structure of the *Methanospirillum hungatei* archaeellum reveals structural features distinct from the bacterial flagellum and type IV pili. *Nat. Microbiol.* **2**, 16222 (2016).
- Daum, B. et al. Structure and in situ organisation of the *Pyrococcus furiosus* archaeellum machinery. *eLife* **6**, e27470 (2017).
- Meshcheryakov, V. A. et al. High-resolution archaeellum structure reveals a conserved metal-binding site. *EMBO reports* **20**, <https://doi.org/10.15252/embr.201846340> (2019).
- Braun, T. et al. Archaeal flagellin combines a bacterial type IV pilin domain with an Ig-like domain. *Proc. Natl Acad. Sci. USA* **113**, 10352–10357 (2016).
- Wang, F. et al. An extensively glycosylated archaeal pilus survives extreme conditions. *Nat. Microbiol.* **4**, 1401–1410 (2019).
- Baquero, D. P. C. P. et al. New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *ISME J.* **14**, 1821–1833 (2020).
- Wang, F. B. et al. M. Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. *Virus Evolution* <https://doi.org/10.1093/ve/veaa023> (2020).
- Egelman, E. H. Ambiguities in helical reconstruction. *eLife* **3**, e04969 (2014).
- Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D. Struct. Biol.* **74**, 814–840 (2018).
- Egelman, E. H. Cryo-EM of bacterial pili and archaeal flagellar filaments. *Curr. Opin. Struct. Biol.* **46**, 31–37 (2017).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
- Reardon, P. N. & Mueller, K. T. Structure of the type IVa major pilin from the electrically conductive bacterial nanowires of *Geobacter sulfurreducens*. *J. Biol. Chem.* **288**, 29260–29266 (2013).
- Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
- Kuhlbrandt, W. Biochemistry The resolution revolution. *Science* **343**, 1443–1444 (2014).
- Egelman, E. H. The current revolution in Cryo-EM. *Biophys. J.* **110**, 1008–1012 (2016).
- Bai, X. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
- Yuan, S. et al. Cryo-EM structure of a herpesvirus capsid at 3.1 Å. *Science* **360**, <https://doi.org/10.1126/science.aao7283> (2018).
- Fang, Q. et al. Near-atomic structure of a giant virus. *Nat. Commun.* **10**, 388 (2019).
- Wang, F. et al. A packing for A-form DNA in an icosahedral virus. *Proc. Natl Acad. Sci. USA* **116**, 22591–22597 (2019).
- Veesler, D. et al. Atomic structure of the 75 MDa extremophile *Sulfolobus turreted* icosahedral virus determined by CryoEM and X-ray crystallography. *Proc. Natl Acad. Sci. USA* **110**, 5504–5509 (2013).
- Kandiba, L. & Eichler, J. Archaeal S-layer glycoproteins: post-translational modification in the face of extremes. *Front Microbiol.* **5**, 661 (2014).
- Klingl, A. & Pickl, C., & Flechler, J. Archaeal cell walls. *Subcell. Biochem.* **92**, 471–493 (2019).
- Ellison, C. K. et al. Obstruction of pilus retraction stimulates bacterial surface sensing. *Science* **358**, 535–538 (2017).
- Zeng, Z., Liu, X. L., Wei, J. H., Summons, R. E. & Welander, P. V. Calditol-linked membrane lipids are required for acid tolerance in *Sulfolobus acidocaldarius*. *Proc. Natl Acad. Sci. USA* **115**, 12932–12937 (2018).
- Beeby, M. et al. The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.* **3**, e309 (2005).
- Krupovic, M., White, M. F., Forterre, P. & Prangishvili, D. Postcards from the edge: structural genomics of archaeal viruses. *Adv. virus Res.* **82**, 33–62 (2012).
- Menon, S. K. et al. Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. *Virology* **376**, 270–278 (2008).
- Hartung, S. et al. Ultrahigh resolution and full-length pilin structures with insights for filament assembly, pathogenic functions, and vaccine potential. *J. Biol. Chem.* **286**, 44254–44265 (2011).
- Sheppard, D. et al. The major subunit of widespread competence (pseudo)pili exhibits a novel and conserved type IV pilin fold. Preprint at <https://doi.org/10.1101/2020.02.12.945410> (2020).
- Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
- Dohmen, E., Klasberg, S., Bornberg-Bauer, E., Perrey, S. & Kemena, C. The modular nature of protein evolution: domain rearrangement rates across eukaryotic life. *BMC Evol. Biol.* **20**, 30 (2020).
- Moore, A. D., Bjorklund, A. K., Ekman, D., Bornberg-Bauer, E. & Elofsson, A. Arrangements in the modular evolution of proteins. *Trends Biochem. Sci.* **33**, 444–451 (2008).
- Rensen, E. I. et al. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proc. Natl Acad. Sci. USA* **113**, 2478–2483 (2016).
- Quemin, E. R. et al. *Sulfolobus* spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J. Virol.* **89**, 11681–11691 (2015).
- Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
- Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
- Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
- Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
- Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
- He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
- Petersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 213–221 (2010).
- Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
- Song, Y. et al. High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).

67. Poirot, O., O'Toole, E. & Notredame, C. Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res.* **31**, 3503–3506 (2003).
68. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 (2004).

## Acknowledgements

This research was, in part, aided by the National Cancer Institute's National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research under contract HSSN261200800001E. Cryo-EM imaging was also conducted at the Molecular Electron Microscopy Core facility at the University of Virginia, which is supported by the School of Medicine and built with NIH grant G20-RR31199. This work was supported by NIH grant R35GM122510 (to E.H.E.). M.K. was supported by l'Agence Nationale de la Recherche (France) project ENVIRA. D.P.B. is part of the Pasteur – Paris University (PPU) International PhD Program, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665807. We are also grateful to the Ultrastructural BioImaging (UTechS UBI) unit of Institut Pasteur for access to electron microscopes.

## Author contributions

D.P.B. prepared all of the samples; F.W. collected the cryo-EM data; F.W., L.B. and E.H.E. performed image processing; F.W. did the structural modeling; F.W., Z.S. and M.K. prepared figures; F.W., M.K. and E.H.E. wrote the manuscript; E.H.E., D.P. and M.K. conceived the study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-17268-4>.

**Correspondence** and requests for materials should be addressed to M.K. or E.H.E.

**Peer review information** *Nature Communications* thanks Han Remaut, Matthias Wolf and the other, anonymous, reviewer(s) 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.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020