Deep conservation of histone variants in Thermococcales archaea

Kathryn M Stevens[1,2], Antoine Hocher[1,2], Tobias Warnecke[1,2*]

[1]Medical Research Council London Institute of Medical Sciences, London, United Kingdom
[2]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom

*corresponding author: tobias.warnecke@lms.mrc.ac.uk

Abstract

Histones are ubiquitous in eukaryotes where they assemble into nucleosomes, binding and wrapping DNA to form chromatin. One process to modify chromatin and regulate DNA accessibility is the replacement of histones in the nucleosome with paralogous variants. Histones are also present in archaea but whether and how histone variants contribute to the generation of different physiologically relevant chromatin states in these organisms remains largely unknown. Conservation of paralogs with distinct properties can provide *prima facie* evidence for defined functional roles. We recently revealed deep conservation of histone paralogs with different properties in the Methanobacteriales, but little is known experimentally about these histones. In contrast, the two histones of the model archaeon *Thermococcus kodakarensis,* HTkA and HTkB, have been examined in some depth, both *in vitro* and *in vivo*. HTkA and HTkB exhibit distinct DNA-binding behaviours and elicit unique transcriptional responses when deleted. Here, we consider the evolution of HTkA/B and their orthologs across the order Thermococcales. We find histones with signature HTkA- and HTkB-like properties to be present in almost all Thermococcales genomes. Phylogenetic analysis indicates the presence of one HTkA- and one HTkB-like histone in the ancestor of Thermococcales and long-term maintenance of these two paralogs throughout Thermococcales diversification. Our results support the notion that archaea and eukaryotes have convergently evolved histone variants that carry out distinct adaptive functions. Intriguingly, we also detect more highly diverged histone-fold proteins, related to those found in some bacteria, in several Thermococcales genomes. The functions of these bacteria-type histones remain unknown, but structural modelling suggests that they can form heterodimers with HTkA/B-like histones.

Significance statement

Histones are key components of chromatin in eukaryotes and many archaea. In eukaryotes, histone variants exist that play defined roles in cellular function and development. Some of these variants are highly conserved and can date back to the last common ancestor of eukaryotes. Archaea also frequently encode multiple sequence-divergent histones but whether these play distinct functional roles that are conserved through evolution remains largely unknown. Here, using phylogenetic tools, we establish the existence of histone variants with different properties in the order Thermococcales, which have been conserved for hundreds of millions of years. Our work provides additional evidence for histone variants in archaea and that these have evolved in parallel to eukaryotes to mediate flexible, adaptive chromatin states.

Introduction

The ability of eukaryotic cells to respond to environmental change and regulate transcription relies on dynamic control of DNA accessibility through chromatin alterations. This involves many different processes, including the addition/removal of histone modifications and the exchange of histone proteins for paralogous variants. Such variants can modify structural properties of the nucleosome or change how it interacts with its binding partners (Talbert & Henikoff 2010; Martire & Banaszynski 2020; Henikoff & Smith 2015). For example, macroH2A has a large C-terminal domain and precipitates transcriptional repression (Martire & Banaszynski 2020; Bönisch & Hake 2012) while cenH3, a fast-evolving H3 variant, is specifically localised to centromeres and involved in chromosome segregation (Palmer et al. 1991; Talbert & Henikoff 2010). Importantly, significant functional changes can come from small differences in sequence. H3.3, for example, is deposited in a replication-independent manner in actively transcribed regions of the genome (Talbert & Henikoff 2010) and important for mammalian development (Sitbon et al. 2020; Jang et al. 2015), but differs from its paralog H3.1 by only five amino acid.

Histones are not exclusive to eukaryotes. Archaeal histone proteins, first discovered in *Methanothermus fervidus* (Starich et al. 1996; Sandman et al. 1990), have since been identified in diverse archaeal lineages (Henneman et al. 2018; Hocher et al. 2021) and are often highly expressed (Hocher et al. 2021). Eukaryotic and archaeal histones share a conserved histone fold (HF) domain, form dimers and tetramers that are structurally very similar, and bind DNA non-specifically, albeit with a preference for more bendable sequences (Rojec et al. 2019; Nalabothula et al. 2013; Bailey et al. 2000; Decanniere et al. 2000; Mattiroli et al. 2017). Unlike eukaryotic histones, almost all archaeal histones lack long terminal extensions ("tails") (Henneman et al. 2018). In at least some instances, archaeal histones have the capacity to form homo- as well as heterodimers and to assemble into long oligomeric structures on DNA (Mattiroli et al. 2017). These extended complexes (Bowerman et al. 2021) can, in theory, consist of different histone paralogs, providing opportunities for chromatin state modulation through the exchange of histones with different properties (Stevens et al. 2020). In fact, many archaea encode two or more sequence-divergent histone paralogs, but whether these paralogs have defined functional roles akin to eukaryotic histone variants, and whether their expression and assembly change dynamically to mediate adaptive chromatin states, remains poorly understood.

What we do know from prior experimental work is that archaeal histone paralogs are more than mere copy number variants. The two histones of *M. fervidus* (HMfA, HMfB), for example, display differences in DNA binding affinity (Bailey et al. 2002). Compared to HMfA, recombinant HMfB also induces more positive supercoiling upon binding to plasmid DNA and forms a more compact complex as inferred from gel-shift and tethered particle motion experiments (Sandman et al. 1994; Henneman et al. 2021). There are also differences between HMfA and HMfB in their relative expression during the growth cycle: in early exponential phase, HMfA is more highly expressed than HMfB, expression of which increases towards stationary phase to reach an almost equal ratio between the two (Sandman et al. 1994). The different properties of *M. fervidus* histones are consistent with the hypothesis that the two paralogs may have distinct functions in nucleoid/chromatin biology, but whether the properties are physiologically relevant and affect organismal fitness remains to be addressed experimentally.

Recently, we considered this question using an evolutionary approach. We identified histone paralogs in the order Methanobacteriales that exhibit distinct structural properties and have been maintained over hundreds of millions of years (Stevens et al. 2020), indicative of the importance of each individual paralog for fitness. Structural modelling identified histone variants that prevent stable tetramerization and might act as *capstones* that limit further extension when incorporated into a histone oligomer, providing a potential pathway to dynamically alter chromatin state. Are the Methanobacteriales unique or are there other clades of archaea with histone paralogs that have been maintained over long periods of time? And do these paralogs also show conserved and distinct structural properties?

Here, we consider archaea in the order Thermococcales, which includes the model archaea *Pyrococcus furiosus* and *Pyrococcus abyssi* as well as *T. kodakarensis*, which has served as a model species for the *in vivo* study of archaeal histones. Thanks to the efforts of Santangelo and co-workers in particular, its two histones – HTkA (TK1413) and HTkB (TK2289) – are arguably the best characterized paralogs *in vivo*. Similar to HMfA and HMfB in *M. fervidus*, HTkA and HTkB can assemble into long oligomers both *in vitro* and *in vivo* (Mattiroli et al. 2017; Sanders et al. 2021; Bowerman et al. 2021; Maruyama et al. 2013). The two histones differ from one another at 11 out of 67 residues (84% identity) and have several distinct properties. HTkA is the more highly expressed paralog, at least in exponential phase, where it makes up 1.1% of the proteome compared to 0.66% for HTkB (Hocher et al. 2021). Together, they are abundant enough to coat the entire *T. kodakarensis* genome (Sanders et al. 2019). HTkB has been shown to bind to DNA more strongly than HTkA and to form more compact complexes, which show faster migration during agarose gel electrophoresis (Higashibata et al. 1999). Deletion of each histone individually results in overlapping but distinct perturbations of the transcriptome (Čuboňová et al. 2012; Sanders et al. 2021). Notably, HTkB-deficient cells exhibit reduced growth, possibly due to changes in the expression, not seen in strains lacking HTkA, of genes that encode translation factors and ribosomal proteins (Čuboňová et al. 2012). Deletion of *htkA* but not *htkB* leads to downregulation of hypothetical membrane proteins and prevents transformation of *T. kodakarensis,* suggesting HTkA alone plays a critical role in DNA uptake and/or integration (Čuboňová et al. 2012).

In this study, we show that histone paralogs with HTkA- and HTkB-like properties are present across the Thermococcales, including Thermococcus, Pyrococcus, and Palaeococus *spp*.. We use structural modelling to show that, in most Thermococcales, HTkB-like histones are predicted to exhibit stronger DNA binding than those with HTkA-like properties. Phylogenetic analysis reveals that HTkA-like histones share a common ancestor to the exclusion of HTkB-like histones and *vice versa*, suggesting that the last common ancestor of the Thermococcales already encoded an HTkA-like and an HTkB-like histone, each of which has been maintained throughout the diversification of this clade for (very) approximately 750 million years (Wolfe & Fournier 2018). The long-term preservation of these two paralogs across the order Thermococcales supports the notion that HTkA/B in *T. kodakarensis* (and their orthologs in other Thermococcales) make unique contributions to genome function and fitness. These findings add further evidence that histone variants are widespread in archaea, evolving in parallel to those in eukaryotes.

Intriguingly, many Thermococcales archaea encode additional types of histone-fold proteins that are similar to histone-fold proteins found in some bacteria (Alva & Lupas 2019). One of these consists of an end-to-end duplication of the histone fold and is rarely found in archaea outside the Thermococcales. Their physiological roles remain unknown, but structural modelling suggest that they are able to heterodimerize with HTkA/B and might therefore further diversify histone-based chromatin states in these archaea.

Results and Discussion

Almost all Thermococcales have histone paralogs with HTkA- and HTkB-like properties

To identify putative histone proteins across the Thermococcales, we scanned 61 predicted proteomes using HMM models and iterative jackhmmer searches (see Methods). Histones with a single histone-fold domain (similar to archaeal HMf-like histones) were found in all genomes (Fig 1c, Table S1). We also recovered putative histone-fold proteins similar to those found in some bacteria (Alva & Lupas 2019), which we discuss further below. A principal component analysis of the HMf-like archaeal histones based on their amino acid properties and isoelectric points (see Methods) suggests that histones can be assigned to one of two groups. One of these groups contains HTkA, the other HTkB (Fig 1a). This is consistent with a previous classification effort that also recovered two major groups of Thermococcales histones (Henneman 2019). Amino acid identities at several residues along the histone fold differ systematically between groups and are diagnostic of group membership. For example, tyrosine is always found at position 35 (Y35) in HTkA-like histones whereas HTkB-like histones have a positively charged lysine (K, 60 out of 62) or histidine (H, 2 out of 62). Similarly, glutamic acid at position 18 (E18) is present in 59 out of the 61 HTkA-like but none of the HTkB-like histones (Fig 1b).

For some of these residues, we know from prior *in vitro* studies – as well as structural modelling – that amino acid identity can affect specific histone properties (Stevens et al. 2020; Soares et al. 2000). For example, substituting Y for K at residue 35 (the amino acids seen in HTKA- and HTKB-like histones, respectively), increases the stability of recombinant histone HFoB from *Methanobacterium formicicum* (Li et al. 2000). In addition, evidence from mass spectrometry indicates that K35 in HTkB from *T. kodakarensis* and *Thermococcus gammatolerans* is acetylated *in vivo* (Alpha-Bazin et al. 2021) although stoichiometry and functional significance of this modification remain to be determined. A tyrosine at the same position in HTkA removes the potential for acetylation. E18 in HMfB forms an intermolecular salt bridge with K53, which helps to stabilise the interaction between monomers in the histone dimer (Sandman et al. 2001; Decanniere et al. 2000). Mutating E18 to proline does not alter DNA binding (Soares et al. 2000) but loss of the intermonomer salt bridge may result in less rigid dimer structures. Finally, having leucine (L) or phenylalanine (F) at residue 46 has no obvious effect on DNA binding in HMfA/B, but the residue, located at the interface between dimers, is important for tetramer formation (Soares et al. 2000; Marc et al. 2002).

We considered how amino acid differences between HTkA- and HTkB-like histones affect two key aspects of the histone-DNA complex: DNA affinity and tetramerization strength, a proxy for tetramer stability. Using a structural modelling approach, we find that predicted DNA binding for HTkB-like paralogs is, in most cases, stronger than for HTkA-like paralogs (Fig 2, see Methods). This is in line with experimental findings that HTkB binds to DNA more tightly and forms a more compact complex with DNA than HTkA (Higashibata et al. 1999). In contrast, predicted tetramerization strength does not strongly discriminate HTkA- from HTkB-like histones (Fig 2). We also considered residues (K14, G17, K26, E30, E34, Q48, E58, K61, K65) that were previously suggested to be important for stacking interactions for either HTkA or HTkB, in the context of longer oligomers (Mattiroli et al. 2017; Henneman et al. 2018, 2021). None of these residues differ substantially between HTkA- and HTkB-like histones, with the exception of Q48 (HTkB) (Henneman et al. 2018). We note that Q48 is relatively uncommon in HTkB-like histones, and that *T. kodakarensis* HTkB may therefore form more stable oligomeric complexes than the HTkB-like histones of most of its cousins.

HtkA- and HTkB-like histones form ancient paralogous groups

Almost all Thermococcales have both an HTkA-like and an HTkB-like histone (Fig 1c). This is consistent with (but not sufficient to demonstrate) ancient paralogy. To unravel the evolutionary history of Thermoccoccales histones, we used RaxML-NG (Kozlov et al. 2019) to build phylogenetic trees of all 123 HMf-like histones found across the 61 genomes in our analysis (see Methods). We find that HTkA-like and HTkB-like paralogs neatly separate into two groups defined by their position on the tree (Fig 1d). This pattern of separation indicates that one HTkA- and one HTkB-like histone were present in the last common ancestor of Thermococcales. We detect only a small number of lineage-specific duplications and deletions and find no evidence of rampant horizontal gene transfer (Fig 1c). The observation that both paralogs have been maintained along divergent Thermococcales lineages strongly suggests that at least some of the amino acid differences between them are functionally important and under selection. Along with our recent report of ancient histone paralogs in the Methanobacteriales (Stevens et al. 2020), this finding provides further evidence that histone variants exist in archaea, evolving in parallel to those in eukaryotes. Note that, at present, we have no convincing evidence that histone paralogs in the Thermococcales and those found in the Methanobacteriales arose from the same ancient duplication event.

Some Thermococcales encode histone-fold proteins similar to those found in bacteria

Our survey also revealed that, alongside the HTkA/B-like histones, many Thermococcales genomes encode histone-fold proteins similar to those found in some bacteria (Fig 1c, Table S1), which harbour either a single or two (pseudodimeric) histone fold domains (Alva & Lupas 2019). We will refer to these as bacteria-type singlets and doublets, respectively. Both types are, on average, less well conserved than HTkA/B-like histones (Fig 3a) and their distribution across the Thermococcales is noticeably patchier (Fig 1c, Fig 3f/g). Neither type is present in the closest sister clades (Methanofastidiosa, Theionarchaea).

The bacteria-type singlet is confined to a monophyletic group that comprises some (but not all) *Thermococcus* spp (Fig 1c, Fig 3f). It is present in all these species in the same highly conserved syntenic context, suggesting a single origin (Fig 3f). Following acquisition, the histone has been maintained along almost all lineages. There is only a single loss event (a clean deletion) in the branch leading to *Thermococcus piezophilus* (Fig 3f). Conserved synteny is also consistent with a single, earlier origin for the bacteria-type doublet, with multiple subsequent losses (Fig 1c, Fig 3g). Based on the branching patterns and different syntenic context, the bacteria-type histones in *Pyrococcus furiosus* have likely been acquired secondarily from an extant or ancient *Thermococcus* species.

Bacteria-type doublets differ considerably in sequence from doublet histones previously described in haloarchaea and *Methanopyrus kandleri* (Fig 3e). Outside the Thermococcales, we only detect additional bacteria-type doublets in some (hyper)thermophilic Methanocaldococcus and Archaeoglobus species, but never their mesophilic relatives, suggestive of horizontal gene transfer in the high-temperature niche.

Bacteria-type histone-fold proteins have only recently been recognized and await functional characterization. The only functional data we have at present comes from transcriptome/proteome profiling. Consistent with lower sequence-level conservation (Fig 3a), the relative expression levels of these genes in *T. kodakarensis* (singlet: TK1040; doublet: TK0750) are lower than those of HTkA/B-like histones at both the transcript and protein level (Fig 3b) (Sas-Chen et al. 2020; Jäger et al. 2014). Together, they make up 0.37% of the measured exponential-phase proteome compared to 0.66% for HTkB and 1.1% for HTkA. TK1040 was previously identified in *T. kodakarensis* chromatin fractions (Maruyama et al. 2011), suggesting a (direct or indirect) association with DNA. However, the same study estimated that less than 1% of the amount of chromatin-associated proteins were attributable to TK1040. We therefore consider it unlikely that these histones are global organizers of DNA similar to HTkA/B-like histones, but might modulate chromatin state, either locally or globally, in response to environmental change. In both *P. furiosus* and *T. kodakarensis,* the bacterial-type doublets are under the control of the heat shock regulator Phr (encoded by PF1790 and TK2291, respectively) and upregulated upon Phr deletion, suggesting a potential role in response to heat shock in these archaea (Kanai et al. 2010; Keese et al. 2010). The *T. kodakarensis* doublet is also downregulated at lower temperatures, similar to HTkA/B (Hocher et al. 2021), further consistent with a role in temperature adaptation.

Can these histone fold proteins interact with HTkA/B-like histones? We used AlphaFold (Jumper et al. 2021) to predict the structure of combinations of HTkA, HTkB and the bacterial HF singlet from *T. kodakarensis*. Using this approach, all three are predicted to form homodimers and, as expected, HTkA and HTkB form a stable heterodimer. When a combination of either HTkA or HTkB and the bacterial singlet are used, Alphafold also predicts that these will form a heterodimer (Fig 3c). The presence of non-HMf-like HF proteins in Thermococcales genomes adds to the potential functional diversity of histone-based chromatin in these species and may dynamically alter DNA accessibility at different stages of cell growth or in response to environmental challenges. Further experimental investigation is required, however, before meaningful conclusions can be drawn in this regard, including whether they do interact, both structurally and functionally, with the HTkA/B-like histones.

Methods

Identification of histones in Thermococcales genomes

Protein sets, genomes and GFF files for all available genomes of class Thermococci were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/assembly) using taxid 183968 [accessed on 2021-05-27]. Genomes not present in the GTDB tree (Parks et al. 2021) (https://gtdb.ecogenomic.org, accessed 2021-08-01, see below) were removed. Two species which were annotated as Thermococci in NCBI but branched outside the main group on the GTDB tree were removed from the analysis, leaving a final set of 61 genomes, all from the order Thermococcales. Protein sequences were predicted using Prodigal v2.6.3 (Hyatt et al. 2010) where not provided through GenBank. Histone proteins were extracted from the protein sets through HMM searches using HMMER v3.3.1 (hmmsearch --noali) (Eddy 2011; Finn et al. 2011) using Pfam models CBFD_NFYB_HMF and DUF1931 (Finn et al. 2014) as well as a Jackhmmer searches using the singlet and doublet histones from bacteria as a seed used by others (Alva & Lupas 2019). Bacteria-type histones hit in the initial search were used to build an HMM model and the Thermococcales protein set was re-searched using HMMER v3.3.1 as above (hmmsearch --noali). Some proteins incorrectly identified as histones at this stage were manually filtered out.

Classification of Thermococcales histones into HtkA-like and HtkB-like groups

HMf-like histones in Thermococcales downloaded from GenBank [accessed on 2021-05-27] (see above) were aligned using MAFFT (Katoh & Standley 2013) (–localpair --maxiterate 1000). Histones were clustered based on amino acid composition of their peptide sequences using AAStats from the R package Seqinr (Charif & Lobry 2007). Histones which clustered with HTkA and HTkB were assigned HTkA- or HTkB-like status, respectively (see Fig 1a). 20 maximum likelihood phylogenetic trees were built using Raxml-NG (Kozlov et al. 2019) with the LG+G4 model of evolution as suggested by ModelTest-NG (Darriba et al. 2020). The unrooted best maximum likelihood tree is shown. All trees were plotted using iTOL (Letunic & Bork 2019). Orthologous genes in the genomic neighbourhood of each histone were highlighted on the tree using Genespy as best reciprocal hits (Garcia et al. 2019). Orthologs in the histone neighbourhood were identified by performing reciprocal best hits for each genome against *T. kodakarensis* using BLAST (Altschul et al. 1990), retaining those which have a similarity score of >40% and are within 20% length of one another (Rocha 2006). Note that the reciprocal best hit approach was only applied to identify putative orthologs in the neighbourhood of histone genes. The histones themselves were identified using HMMer searches (see above) and subsequent analyses not restricted to reciprocal best hits of HTkA/B. To generate sequence logos, histones were aligned using MAFFT (Katoh & Standley 2013) (–localpair --maxiterate 1000) and visualized using ggseqlogo in R (Wagih 2017).

Predicted DNA binding and tetramerisation

Predicted DNA-binding and interaction (tetramerization) strength between dimers for Thermococcales species was computed as in (Stevens et al. 2020). In brief, sequences were aligned to HMfB and substitutions were mapped onto a tetrameric model of HMfB (extracted from PDB structure 5t5k) using FoldX (Schymkowitz et al. 2005) to generate models of homotetramers with DNA. Structures were then energy-minimised using AmberTools (Maier et al. 2015) and binding affinity was calculated using an MMPBSA approach with the ff14SB forcefield (Miller et al. 2012). $\Delta\Delta G$ was calculated relative to HMfB. The mean value for five replicates is shown for each model.

Species tree

The archaeal species tree was downloaded from GTDB (https://gtdb.ecogenomic.org) on 2021-08-01.

Expression data

Expression data for *T. kodakarensis* was obtained from primary sources and NCBI's Gene Expression Omnibus (GEO) (Barrett et al. 2013). Proteomics data from (Sas-Chen et al. 2020) was processed as in Hocher et al. (2021). Protein abundance at 85°C is shown. Transcript abundance data was obtained from (Jäger et al. 2014) and is shown as transcripts per million (TPM).

Bacteria-type singlet histone fold and HTkA/B structure prediction

The AlphaFold v2.0 (Jumper et al. 2021) collab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb accessed 18-08-21) was used to predict the structure of HTkA, HTkB and the bacterial singlet HF protein (TK1040) as homodimers and all heterodimer combinations. The MSA method used was jackhmmer, and models were ranked by PTMscore. The top model is shown for all homodimers and heterodimers. Images shown were generated using UCSF, ChimeraX developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health (R01-GM129325) and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (Pettersen et al. 2021).

Identification of doublet histones in archaea

The genomes, protein sets and GFF files for a balanced set of archaea species was downloaded on 2021-05-21 from GenBank (https://www.ncbi.nlm.nih.gov/assembly) using taxid 2157, and processed as above (see: *Identification of histones in Thermococcales genomes*). Doublet histones, including doublets from Halobacteria (Dulmage et al. 2015) and HMk in *Methanopyrus kandleri* (Fahrner et al. 2001; Slesarev et al. 1998) were identified by their position on a CDS-level tree of archaeal histones, and by length. This larger tree was built from 20 ML trees using Raxml-NG (Kozlov et al. 2019). Trees for the doublet histones were then built as previously described (see: *Classification of Thermococcales histones into HtkA-like and HtkB-like groups*) and all figures plotted using iTOL (Letunic & Bork 2019). Annotation for the orthologous genes in the genomic neighbourhood was generated using

GeneSpy (Garcia et al. 2019). Clades with doublet histones are annotated on a tree of archaea adapted from (Borrel et al. 2020) to include Hodarchaeota (Liu et al. 2021).

Percentage identity of HTkA-/HTkB-like and bacteria-type histones

HTkA-like, HTkB-like and bacteria-type histones were aligned using MAFFT (Katoh & Standley 2013) (–localpair --maxiterate 1000) separately for species containing either bacteria-type doublet or bacteria-type singlet histones. For each histone type, pairwise sequence identity was calculated using seqidentity from the R package bio3d (Grant et al. 2006).

Data Availability Statement

All data underlying were derived from sources in the public domain: GenBank (https://www.ncbi.nlm.nih.gov/assembly); GTDB (https://gtdb.ecogenomic.org); and data from original publications cited throughout the manuscript.

References

Alpha-Bazin B et al. 2021. Lysine-specific acetylated proteome from the archaeon Thermococcus gammatolerans reveals the presence of acetylated histones. J. Proteomics. 232:104044. doi: 10.1016/j.jprot.2020.104044.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410. doi: 10.1016/S0022-2836(05)80360-2.

Alva V, Lupas AN. 2019. Histones Predate the Split Between Bacteria and Archaea. Bioinformatics. 35:2349–2353. doi: 10.1093/bioinformatics/bty1000.

Bailey KA, Marc F, Sandman K, Reeve JN. 2002. Both DNA and histone fold sequences contribute to archaeal nucleosome stability. J. Biol. Chem. 277:9293–9301. doi: 10.1074/jbc.M110029200.

Bailey KA, Pereira SL, Widom J, Reeve JN. 2000. Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. J. Mol. Biol. 303:25–34. doi: 10.1006/jmbi.2000.4128.

Barrett T et al. 2013. NCBI GEO: Archive for functional genomics data sets - Update.

Nucleic Acids Res. 41:991–995. doi: 10.1093/nar/gks1193.
Bönisch C, Hake SB. 2012. Histone H2A variants in nucleosomes and chromatin: More or less stable? Nucleic Acids Res. 40:10719–10741. doi: 10.1093/nar/gks865.
Borrel G, Brugère JF, Gribaldo S, Schmitz RA, Moissl-Eichinger C. 2020. The host-associated archaeome. Nat. Rev. Microbiol. doi: 10.1038/s41579-020-0407-y.
Bowerman S, Wereszczynski J, Luger K. 2021. Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. Elife. 10:e65587. doi: 10.7554/eLife.65587.
Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Structural approaches to sequence evolution: Molecules, networks, populations. Bastolla, U, Porto, M, Roman, HE, & Vendruscolo, M, editors. Springer Verlag: New York pp. 207–232.
Čuboňová L et al. 2012. An archaeal histone is required for transformation of Thermococcus kodakarensis. J. Bacteriol. 194:6864–6874. doi: 10.1128/JB.01523-12.
Darriba Di et al. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol. Biol. Evol. 37:291–294. doi: 10.1093/molbev/msz189.
Decanniere K, Babu AM, Sandman K, Reeve JN, Heinemann U. 2000. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon Methanothermus fervidus. J. Mol. Biol. 303:35–47. doi: 10.1006/jmbi.2000.4104.
Dulmage KA, Todor H, Schmid AK. 2015. Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. MBio. 6. doi: 10.1128/mbio.00649-15.
Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7:e1002195. doi: 10.1371/journal.pcbi.1002195.
Fahrner RL, Cascio D, Lake JA, Slesarev A. 2001. An ancestral nuclear protein assembly: Crystal structure of the Methanopyrus kandleri histone. Protein Sci. 10:2002–2007. doi: 10.1110/ps.10901.
Finn RD et al. 2014. Pfam: The protein families database. Nucleic Acids Res. 42:222–230. doi: 10.1093/nar/gkt1223.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res. 39:29–37. doi: 10.1093/nar/gkr367.
Garcia PS, Jauffrit F, Grangeasse C, Brochier-Armanet C. 2019. GeneSpy, a user-friendly and flexible genomic context visualizer. Bioinformatics. 35:329–331. doi:

10.1093/bioinformatics/bty459.

Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 22:2695–2696. doi: 10.1093/bioinformatics/btl461.

Henikoff S, Smith MM. 2015. Histone Variants and Epigenetics. Cold Spring Harb. Perspect. Biol. 7:a019364. doi: 10.1101/cshperspect.a019364.

Henneman B. 2019. Histone-DNA assemblies in archaea: shaping the genome on the edge of life. Leiden University.

Henneman B et al. 2021. Mechanical and structural properties of archaeal hypernucleosomes. Nucleic Acids Res. 49:4338–4349. doi: 10.1093/nar/gkaa1196.

Henneman B, van Emmerik C, van Ingen H, Dame RT. 2018. Structure and function of archaeal histones. PLoS Genet. 14:e1007582. doi: 10.1371/journal.pgen.1007582.

Higashibata H, Fujiwara S, Takagi M, Imanaka T. 1999. Analysis of DNA compaction profile and intracellular contents of archaeal histones from Pyrococcus kodakaraensis KOD1. Biochem. Biophys. Res. Commun. 258:416–424. doi: 10.1006/bbrc.1999.0533.

Hocher A et al. 2021. Growth temperature is the principal driver of chromatinization in archaea. BioRxiv. doi: https://doi.org/10.1101/2021.07.08.451601.

Hyatt D et al. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11:119. doi: 10.1186/1471-2105-11-119.

Jäger D, Förstner KU, Sharma CM, Santangelo TJ, Reeve JN. 2014. Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. BMC Genomics. 15:684. doi: 10.1186/1471-2164-15-684.

Jang C, Shibata Y, Starmer J, Yee D, Magnuson T. 2015. Histone H3.3 maintains genome integrity during mammalian development. Genes Dev. 29:1377–1392. doi: 10.1101/gad.264150.115.GENES.

Jumper J et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature. 596:583–589. doi: 10.1038/s41586-021-03819-2.

Kanai T, Takedomi S, Fujiwara S, Atomi H, Imanaka T. 2010. Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, Thermococcus kodakaraensis. J. Biochem. 147:361–370. doi: 10.1093/jb/mvp177.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780. doi: 10.1093/molbev/mst010.

Keese AM, Schut GJ, Ouhammouch M, Adams MWW, Thomm M. 2010. Genome-wide identification of targets for the archaeal heat shock regulator Phr by cell-free transcription of genomic DNA. J. Bacteriol. 192:1292–1298. doi: 10.1128/JB.00924-09.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 35:4453–4455. doi: 10.1093/bioinformatics/btz305.

Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 47:256–259. doi: 10.1093/nar/gkz239.

Li WT, Shriver JW, Reeve JN. 2000. Mutational analysis of differences in thermostability between histones from mesophilic and hyperthermophilic archaea. J. Bacteriol. 182:812–817. doi: 10.1128/JB.182.3.812-817.2000.

Liu Y et al. 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature. 593:553–557. doi: 10.1038/s41586-021-03494-3.

Maier JA et al. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 11:3696–3713. doi: 10.1021/acs.jctc.5b00255.

Marc F, Sandman K, Lurz R, Reeve JN. 2002. Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. J. Biol. Chem. 277:30879–30886. doi: 10.1074/jbc.M203674200.

Martire S, Banaszynski LA. 2020. The roles of histone variants in fine-tuning chromatin organization and function. Nat. Rev. Mol. Cell Biol. 21:522–541. doi: 10.1038/s41580-020-0262-8.

Maruyama H et al. 2013. An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. EMBO Rep. 14:711–717. doi: 10.1038/embor.2013.94.

Maruyama H et al. 2011. Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis. Mol. Biol. Cell. 22:386–398. doi: 10.1091/mbc.e10-08-0668.

Mattiroli F et al. 2017. Structure of histone-based chromatin in Archaea. Science. 357:609–612. doi: 10.1126/science.aaj1849.

Miller BR et al. 2012. MMPBSA.py: An efficient program for end-state free energy calculations. J. Chem. Theory Comput. 8:3314–3321. doi: 10.1021/ct300418h.

Nalabothula N et al. 2013. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. BMC Genomics. 14:391. doi: 10.1186/1471-2164-14-391.

Palmer DK, O'Day K, Trong HLE, Charbonneau H, Margolis RL. 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc.

Natl. Acad. Sci. U. S. A. 88:3734–3738. doi: 10.1073/pnas.88.9.3734.

Parks DH et al. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. doi: 10.1093/nar/gkab776.

Pettersen EF et al. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30:70–82. doi: 10.1002/pro.3943.

Rocha EPC. 2006. Inference and analysis of the relative stability of bacterial chromosomes. Mol. Biol. Evol. 23:513–522. doi: 10.1093/molbev/msj052.

Rojec M, Hocher A, Stevens KM, Merkenschlager M, Warnecke T. 2019. Chromatinization of Escherichia coli with archaeal histones. Elife. 8:e49038. doi: 10.7554/eLife.49038.

Sanders TJ et al. 2021. Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis. Front. Microbiol. 12:1071. doi: 10.3389/fmicb.2021.681150.

Sanders TJ et al. 2019. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. Mol. Microbiol. 111:784–797. doi: 10.1111/mmi.14191.

Sandman K, Grayling RA, Dobrinski B, Lurz R, Reeve JN. 1994. Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus. Proc. Natl. Acad. Sci. 91:12624–12628. doi: 10.1073/pnas.91.26.12624.

Sandman K, Krzycki JA, Dobrinski B, Lurz R, Reeve JN. 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histones. Proc. Natl. Acad. Sci. 87:5788–5791. doi: 10.1073/pnas.87.15.5788.

Sandman K, Soares D, Reeve JN. 2001. Molecular components of the archaeal nucleosome. Biochimie. 83:277–281. doi: 10.1016/S0300-9084(00)01208-6.

Sas-Chen A et al. 2020. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature. 583:638–643. doi: 10.1038/s41586-020-2418-2.

Schymkowitz J et al. 2005. The FoldX web server: An online force field. Nucleic Acids Res. 33:W382–W388. doi: 10.1093/nar/gki387.

Sitbon D, Boyarchuk E, Dingli F, Loew D, Almouzni G. 2020. Histone variant H3.3 residue S31 is essential for Xenopus gastrulation regardless of the deposition pathway. Nat. Commun. 11:1256. doi: 10.1038/s41467-020-15084-4.

Slesarev AI, Belova GI, Kozyavkin SA, Lake JA. 1998. Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. Nucleic Acids Res. 26:427–430. doi: 10.1093/nar/26.2.427.

Soares DJ, Sandman K, Reeve JN. 2000. Mutational analysis of archaeal histone-DNA interactions. J. Mol. Biol. 297:39–47. doi: 10.1006/jmbi.2000.3546.

Starich MR, Sandman K, Reeve JN, Summers MF. 1996. NMR structure of HMfB from the hyperthermophile, Methanothermus fervidus, confirms that this archaeal protein is a histone. J. Mol. Biol. 255:187–203. doi: 10.1006/jmbi.1996.0016.

Stevens KM et al. 2020. Histone variants in archaea and the evolution of combinatorial chromatin complexity. Proc. Natl. Acad. Sci. U. S. A. 117:33384–33395. doi: 10.1073/PNAS.2007056117.

Talbert PB, Henikoff S. 2010. Histone variants - ancient wrap artists of the epigenome. Nat. Rev. Mol. Cell Biol. 11:264–275. doi: 10.1038/nrm2861.

Wagih O. 2017. ggseqlogo: A versatile R package for drawing sequence logos. Bioinformatics. 33:3645–3647. doi: 10.1093/bioinformatics/btx469.

Wolfe JM, Fournier GP. 2018. Horizontal gene transfer constrains the timing of methanogen evolution. Nat. Ecol. Evol. 2:897–903. doi: 10.1038/s41559-018-0513-7.

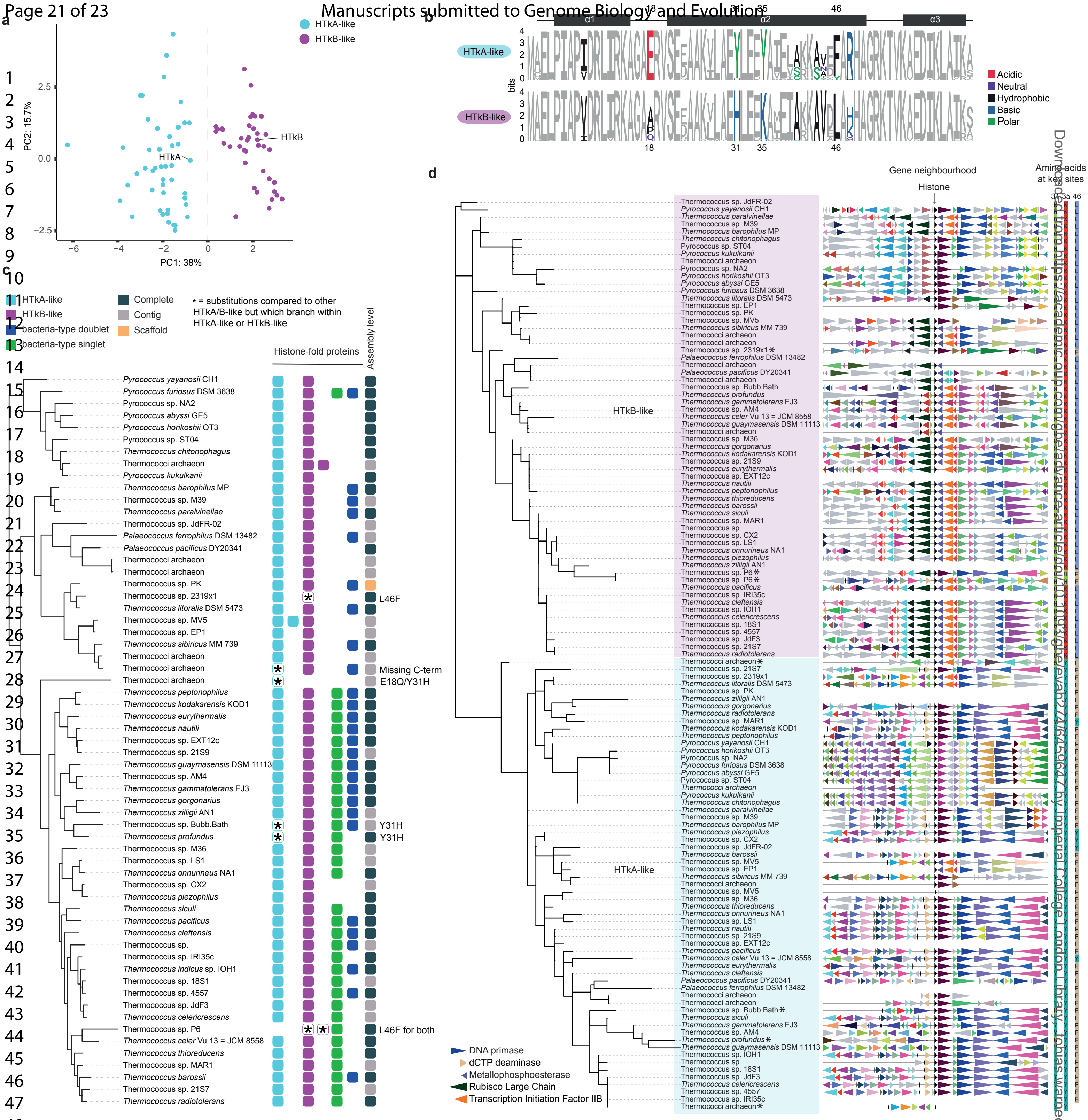
Figure legends

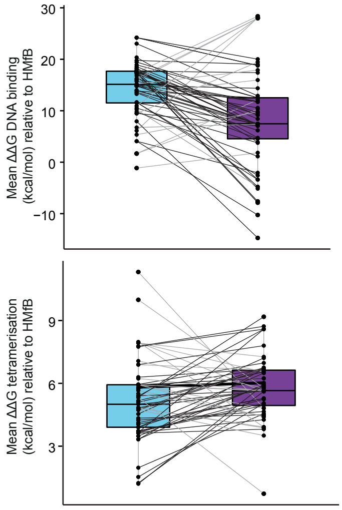
Figure 1. a. Principle component analysis of Thermococcales HMf-like histones based on amino acid properties (AAStats, see Methods). Histones that cluster with either HTkA or HTkB along the first principal component are coloured accordingly (HTkA-like histones in blue, HTkB-like in purple). **b.** Sequence logos showing amino acid composition of HTkA- and HTkB-like histones across 61 Thermococcales. Amino acids that differ substantially between the two groups are coloured. Positions are numbered relative to HMfB from *Methanothermus fervidus* to facilitate comparison with prior studies. **c.** GTDB species tree for all Thermococcales in the dataset indicating presence/absence of histones of a particular type in each genome. Each square represents one histone and is coloured by histone type. **d.** Protein-level phylogenetic tree of all HMf-like Thermococcales histones. Genes in the neighbourhood are coloured to indicate ortholog identity (see Methods). Note that, while the gene neighbourhood is broadly conserved for HTkA- vis-à-vis HTkB-like orthologs, some HTkB-like genes (e.g. in *Thermococcus chitonophagus*) have a 3' neighbourhood normally found for HTkA. This is owing to a large genomic rearrangement event in which HTkB served as the breakpoint (not shown).

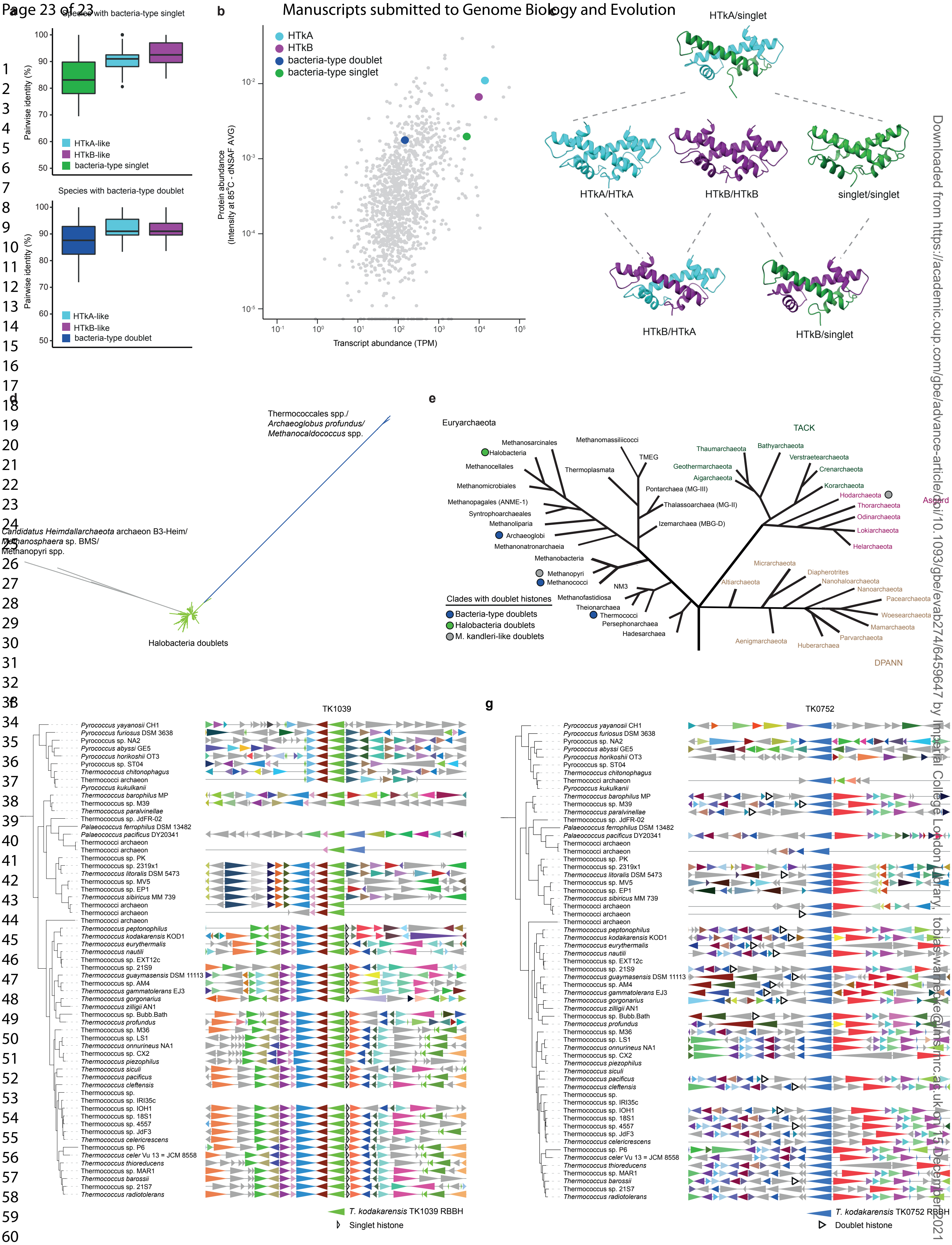
Figure 2. Predicted DNA binding affinity (top) and tetramer stability (bottom) for HTkA/B-

like paralogs. Lines connect paralogs from the same genome. Lines are black when DNA binding affinity is stronger or tetramer interface energy weaker, respectively, for the HTkB-like paralog.

Figure 3. a. Pairwise identity for histones in species with HTkA-like, HTkB-like, and bacteria-type singlet (top) or doublet histones (bottom) across all Thermococcales. **b.** Protein and transcript abundance (see Methods) of genes in *T. kodakarensis.* TPM: transcripts per million. **c.** AlphaFold-predicted homodimeric structures of TK1040 (green), HTkA (light blue) and HTkB (purple) and heterodimers of HTkA and HTkB (light blue/purple), HTkA and the bacteria-type singlet TK1040 (light blue/green), and HTkB and TK1040 (purple/green). **d.** Protein-level tree for doublet histones (containing an end-to-end duplication) in archaea. Green branches contain proteins from Halobacteria, blue branches contain bacteria-type doublet histones in Thermococcales, *Methanocaldococcus* spp. and one *Archaeglobus* sp., grey branches show doublet histones including HMk from *Methanopyrus kandleri* (Fahrner et al. 2001; Slesarev et al. 1998). **e.** Clades with doublet histones highlighted on an tree capturing archaeal diversity, adapted from (Borrel et al. 2020; Stevens et al. 2020) to include the clade for the species *Candidatus Heimdallarchaeota* archaeon B3-Heim, more recently suggested to be a member the Hodarchaeota (Liu et al. 2021). Circles denote clades containing a doublet histone, and are coloured by the type of doublet (see d). **f,g.** Thermococcales GTDB species tree showing syntenic regions for TK1039 (light green, in **f)** and TK0752 (blue, in **g**) commonly found located close to the bacteria-type singlet or doublet histones. Genes are coloured to indicate homology (see Fig 1d/Methods).


a
HTkA-like
HTkB-like
HTkA
HTkB
PC1: 38%
PC2: 15.7%
b
α1
α2
α3
18
31
35
46
bits
HTkA-like
HTkB-like
Acidic
Neutral
Hydrophobic
Basic
Polar
c
HTkA-like
HTkB-like
bacteria-type doublet
bacteria-type singlet
Complete
Contig
Scaffold
* = substitutions compared to other HTkA/B-like but which branch within HTkA-like or HTkB-like
Histone-fold proteins
Assembly level
Pyrococcus yayanosii CH1
Pyrococcus furiosus DSM 3638
Pyrococcus sp. NA2
Pyrococcus abyssi GE5
Pyrococcus horikoshii OT3
Pyrococcus sp. ST04
Thermococcus chitonophagus
Thermococci archaeon
Pyrococcus kukulkanii
Thermococcus barophilus MP
Thermococcus sp. M39
Thermococcus paralvinellae
Thermococcus sp. JdFR-02
Palaeococcus ferrophilus DSM 13482
Palaeococcus pacificus DY20341
Thermococci archaeon
Thermococci archaeon
Thermococcus sp. PK
Thermococcus sp. 2319x1
Thermococcus litoralis DSM 5473
Thermococcus sp. MV5
Thermococcus sp. EP1
Thermococcus sibiricus MM 739
Thermococci archaeon
Thermococci archaeon
Thermococci archaeon
Thermococcus peptonophilus
Thermococcus kodakarensis KOD1
Thermococcus eurythermalis
Thermococcus nautili
Thermococcus sp. EXT12c
Thermococcus sp. 21S9
Thermococcus guaymasensis DSM 11113
Thermococcus sp. AM4
Thermococcus gammatolerans EJ3
Thermococcus gorgonarius
Thermococcus zilligii AN1
Thermococcus sp. Bubb.Bath
Thermococcus profundus
Thermococcus sp. M36
Thermococcus sp. LS1
Thermococcus onnurineus NA1
Thermococcus sp. CX2
Thermococcus piezophilus
Thermococcus siculi
Thermococcus pacificus
Thermococcus cleftensis
Thermococcus sp.
Thermococcus sp. IRI35c
Thermococcus indicus sp. IOH1
Thermococcus sp. 18S1
Thermococcus sp. 4557
Thermococcus sp. JdF3
Thermococcus celericrescens
Thermococcus sp. P6
Thermococcus celer Vu 13 = JCM 8558
Thermococcus thioreducens
Thermococcus sp. MAR1
Thermococcus barossii
Thermococcus sp. 21S7
Thermococcus radiotolerans
L46F
Missing C-term
E18Q/Y31H
Y31H
Y31H
L46F for both
d
Gene neighbourhood
Histone
Amino acids at key sites
31 35 46
HTkB-like
HTkA-like
DNA primase
dCTP deaminase
Metallophosphoesterase
Rubisco Large Chain
Transcription Initiation Factor IIB


Mean ΔΔG DNA binding
(kcal/mol) relative to HMfB
30
20
10
0
−10
Mean ΔΔG tetramerisation
(kcal/mol) relative to HMfB
9
6
3


a
Species with bacteria-type singlet
Pairwise identity (%)
HTkA-like
HTkB-like
bacteria-type singlet
Species with bacteria-type doublet
bacteria-type doublet
b
HTkA
HTkB
bacteria-type doublet
bacteria-type singlet
Protein abundance
(Intensity at 85°C - dNSAF AVG)
Transcript abundance (TPM)
c
HTkA/singlet
HTkA/HTkA
HTkB/HTkB
singlet/singlet
HTkB/HTkA
HTkB/singlet
d
Thermococcales spp./
Archaeoglobus profundus/
Methanocaldococcus spp.
Candidatus Heimdallarchaeota archaeon B3-Heim/
Methanosphaera sp. BMS/
Methanopyri spp.
Halobacteria doublets
e
Euryarchaeota
TACK
Asgard
DPANN
Methanosarcinales
Halobacteria
Methanocellales
Methanomicrobiales
Methanopagales (ANME-1)
Syntrophoarchaeales
Methanoliparia
Archaeoglobi
Methanonatronarchaeia
Methanobacteria
Methanopyri
Methanococci
Methanomassiliicocci
Thermoplasmata
TMEG
Pontarchaea (MG-III)
Thalassoarchaea (MG-II)
Izemarchaea (MBG-D)
NM3
Methanofastidiosa
Theionarchaea
Thermococci
Persephonarchaea
Hadesarchaea
Thaumarchaeota
Bathyarchaeota
Verstraetearchaeota
Geothermarchaeota
Aigarchaeota
Crenarchaeota
Korarchaeota
Hodarchaeota
Thorarchaeota
Odinarchaeota
Lokiarchaeota
Helarchaeota
Micrarchaeota
Diapherotrites
Altiarchaeota
Nanohaloarchaeota
Nanoarchaeota
Pacearchaeota
Woesearchaeota
Mamarchaeota
Parvarchaeota
Aenigmarchaeota
Huberarchaea
Clades with doublet histones
Bacteria-type doublets
Halobacteria doublets
M. kandleri-like doublets
f
TK1039
T. kodakarensis TK1039 RBBH
Singlet histone
g
TK0752
T. kodakarensis TK0752 RBBH
Doublet histone