



# Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in *Thermococcus kodakarensis*

Travis J. Sanders<sup>1</sup>, Fahad Ullah<sup>2</sup>, Alexandra M. Gehring<sup>3†</sup>, Brett W. Burkhardt<sup>1</sup>, Robert L. Vickerman<sup>1</sup>, Sudili Fernando<sup>1</sup>, Andrew F. Gardner<sup>3</sup>, Asa Ben-Hur<sup>2</sup> and Thomas J. Santangelo<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Changyi Zhang,  
Carl R. Woese Institute for Genomic  
Biology, University of Illinois  
at Urbana–Champaign, United States

### Reviewed by:

Remus T. Dame,  
Leiden University, Netherlands  
Zhenfeng Zhang,  
Institute of Microbiology, Chinese  
Academy of Sciences, China

### \*Correspondence:

Thomas J. Santangelo  
thomas.santangelo@colostate.edu

### † Present address:

Alexandra M. Gehring,  
Arrakis Therapeutics, Waltham, MA,  
United States

### Specialty section:

This article was submitted to  
Biology of Archaea,  
a section of the journal  
Frontiers in Microbiology

**Received:** 16 March 2021

**Accepted:** 12 April 2021

**Published:** 13 May 2021

### Citation:

Sanders TJ, Ullah F, Gehring AM,  
Burkhardt BW, Vickerman RL,  
Fernando S, Gardner AF, Ben-Hur A  
and Santangelo TJ (2021) Extended  
Archaeal Histone-Based Chromatin  
Structure Regulates Global Gene  
Expression in *Thermococcus*  
*kodakarensis*.  
*Front. Microbiol.* 12:681150.  
doi: 10.3389/fmicb.2021.681150

<sup>1</sup> Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO, United States,

<sup>2</sup> Department of Computer Science, Colorado State University, Fort Collins, CO, United States, <sup>3</sup> Molecular Enzymology  
Division, New England Biolabs, Inc., Ipswich, MA, United States

Histone proteins compact and organize DNA resulting in a dynamic chromatin architecture impacting DNA accessibility and ultimately gene expression. Eukaryotic chromatin landscapes are structured through histone protein variants, epigenetic marks, the activities of chromatin-remodeling complexes, and post-translational modification of histone proteins. In most Archaea, histone-based chromatin structure is dominated by the helical polymerization of histone proteins wrapping DNA into a repetitive and closely gyred configuration. The formation of the archaeal-histone chromatin-superhelix is a regulatory force of adaptive gene expression and is likely critical for regulation of gene expression in all histone-encoding Archaea. Single amino acid substitutions in archaeal histones that block formation of tightly packed chromatin structures have profound effects on cellular fitness, but the underlying gene expression changes resultant from an altered chromatin landscape have not been resolved. Using the model organism *Thermococcus kodakarensis*, we genetically alter the chromatin landscape and quantify the resultant changes in gene expression, including unanticipated and significant impacts on provirus transcription. Global transcriptome changes resultant from varying chromatin landscapes reveal the regulatory importance of higher-order histone-based chromatin architectures in regulating archaeal gene expression.

**Keywords:** archaea, histone, chromatin, transcriptome, *Thermococcus*, RNA-seq

## INTRODUCTION

Most Archaea encode histone proteins to organize their DNA into a protein:DNA complex known as chromatin (Sandman and Reeve, 2000, 2001, 2006; Peeters et al., 2015; Mattioli et al., 2017; Bhattacharyya et al., 2018; Henneman et al., 2018; Sanders et al., 2019b; Henneman et al., 2020; Stevens et al., 2020; Bowerman et al., 2021; Laursen et al., 2021). The organization and tractability of chromatin structure is essential to regulate adaptive gene expression (Wilkinson et al., 2010; Nalabothula et al., 2013; Mattioli et al., 2017). Archaeal and eukaryotic histone proteins are highly homologous, sharing conserved residues that contact DNA and retain structural elements of the

histone fold (Sandman and Reeve, 2000; Soares et al., 2000; Mattioli et al., 2017). In most cases, however, the extensions or tails common to eukaryotic histones are not present in archaeal histone isoforms (Peeters et al., 2015; Mattioli et al., 2017; Nishida and Oshima, 2017; Bhattacharyya et al., 2018; Henneman et al., 2018; Henneman et al., 2020; Sanders et al., 2019b; Stevens et al., 2020). Unlike their eukaryotic counterparts, archaeal histones can homodimerize and spontaneously oligomerize to form chromatin with a single protein (Decanniere et al., 2000; Mattioli et al., 2017; Bhattacharyya et al., 2018; Henneman et al., 2020; Stevens et al., 2020; Bowerman et al., 2021; Laursen et al., 2021), and thus unlike their eukaryotic counterparts (Luger et al., 1997), archaeal chromatin structures are not defined in size. Continued polymerization of archaeal histone proteins produces a symmetrical superstructure composed of increasing lengths of DNA wrapped around a core of polymerized histone dimers (Mattioli et al., 2017). This extended histone-based chromatin structure, alternatively termed the hypernucleosome (Bhattacharyya et al., 2018; Henneman et al., 2018; Henneman et al., 2020), is the biological form of archaeal chromatin within which gene expression is regulated (Figure 1A).

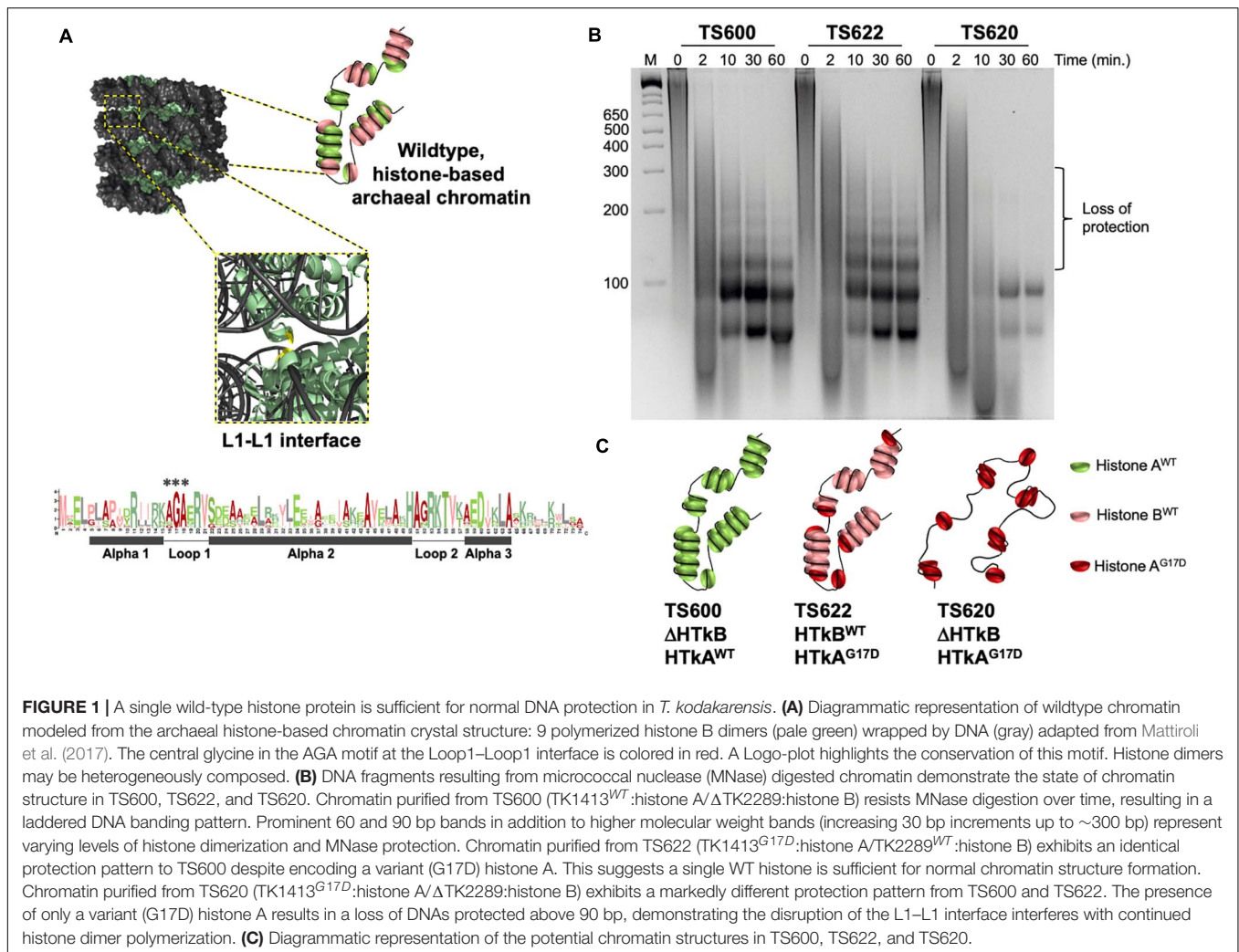
The polymerization of histone dimers in Archaea does not resemble the nucleosome-nucleosome interactions observed in eukaryotes (Luger et al., 1997; Mattioli et al., 2017; Sanders et al., 2019b; Henneman et al., 2020; Stevens et al., 2020; Bowerman et al., 2021). Alignments and analyses of unique archaeal histone sequences reveals the vast majority of archaeal histone proteins retain only the residues comprising the core eukaryotic histone-fold and that many amino acids are highly conserved in positions that form the histone-fold or mediate DNA interactions (Mattioli et al., 2017; Nishida and Oshima, 2017; Zaremba-Niedzwiedzka et al., 2017; Henneman et al., 2020; Stevens et al., 2020). The core histone fold, common to eukaryotic and archaeal histones, consists of three alpha helices ( $\alpha 1$ ,  $\alpha 2$ , and  $\alpha 3$ ) connected through two flexible loops (L1 and L2). A conserved alanine-glycine-alanine (AGA) motif in L1 was identified that was not obviously involved in protein-DNA interactions or in supporting the core histone-fold (Mattioli et al., 2017; Bhattacharyya et al., 2018; Henneman et al., 2018; Henneman et al., 2020; Stevens et al., 2020) (Figure 1B). The crystal structure of archaeal histone-based chromatin provided the first clues to the importance of the highly conserved motif in L1, wherein the small side chains of these nearly invariant amino acids permitted the close proximity of adjacent gyres of DNA within the superhelix at the Loop1-Loop1 (L1-L1) interface. The conservation of these residues was thus hypothesized to support the biologically important, tight-packing of archaeal histone-based chromatin architectures (Mattioli et al., 2017; Bhattacharyya et al., 2018; Henneman et al., 2018; Henneman et al., 2020; Stevens et al., 2020).

Archaeal histone-based chromatin digestion patterns reveal the existence of histone-DNA superstructures that vary in length from protection of  $\sim 60$  bp of DNA by a histone tetramer to protection of  $\sim 300$  bp of DNA by a semi-continuous polymer of  $\sim 10$  symmetrically bound histone dimers. Given the different lengths and proposed dynamic structures of archaeal

histone-based chromatin, we adopt the term archaeal histone-based chromatin polymers (AHCPs) to denote histone:DNA superstructures wherein each additional histone dimer can wrap  $\sim 30$  bp of DNA to extend the AHCP (Nalabothula et al., 2013; Mattioli et al., 2017; Bhattacharyya et al., 2018; Sanders et al., 2019a; Bowerman et al., 2021). While dynamic changes in AHCPs are expected, particularly in species with multiple histone isoforms (Henneman et al., 2020; Stevens et al., 2020), it is also likely that specific DNA sequences and loci are more likely to retain extended AHCPs that are important for regulating gene expression (Čuboňová et al., 2012; Nalabothula et al., 2013; Dulmage et al., 2015). Although the presence of several histone isoforms in many Archaea suggests the potential for variation in superhelix composition and length (Mattioli et al., 2017; Bhattacharyya et al., 2018; Henneman et al., 2018; Sanders et al., 2019b; Henneman et al., 2020; Stevens et al., 2020; Bowerman et al., 2021), the viability of strains with only a single histone demonstrate that even homopolymers permit formation of AHCPs *in vivo* (Sandman and Reeve, 2001, 2006; Santangelo et al., 2009; Čuboňová et al., 2012; Dulmage et al., 2015). The biological importance of AHCPs is supported by the evolutionary retention of the AGA motif in L1 which permits close association of adjacent superhelical gyres. Importantly, allelic substitutions of the central glycine within the L1-AGA motif abrogated formation of extended AHCPs, decreased cellular fitness and resulted in the loss of adaptive gene expression when actively growing cultures were moved to a new metabolic environment (Jäger et al., 2014; Mattioli et al., 2017).

The importance of AHCPs in modulating gene expression suggests unique archaeal gene regulation strategies that take advantage of mechanisms to retain or abolish extended archaeal histone-based chromatin structures (Sanders et al., 2019b; Stevens et al., 2020; Bowerman et al., 2021). To evaluate the normal contribution of AHCPs to gene regulation, we generated strains of the model archaeon *Thermococcus kodakarensis* (Farkas et al., 2013; Gehring et al., 2017; Atomi and Reeve, 2019) wherein genomically encoded histone variants impacted global genomic architecture and quantified the gene expression changes resultant from modified AHCP architectures. *T. kodakarensis* normally encodes two closely related histone isoforms, termed Histone A (HTkA) and Histone B (HTkB), but strains encoding only a single histone variant are viable (Čuboňová et al., 2012; Mattioli et al., 2017; Sanders et al., 2019b). To evaluate the consequences of altering AHCPs in archaeal cells, strains of *T. kodakarensis* were constructed to encode only HTkA in WT form (HTkA<sup>WT</sup>;  $\Delta$ HTkB), only HTkA with a single amino acid substitution G17D (HTkA<sup>G17D</sup>;  $\Delta$ HTkB), or retain HTkB in WT form and express HTkA<sup>G17D</sup> (HTkA<sup>G17D</sup>;  $\Delta$ HTkB<sup>WT</sup>). Substitution of just a single residue, G17, is sufficient to disrupt AHCP formation *in vivo* beyond  $\sim 90$  bp by inhibiting the close association of adjacent gyres of AHCPs (Bhattacharyya et al., 2018). We report here that retention of a single WT histone variant is sufficient to maintain extended AHCPs in archaeal cells, but extended AHCPs are abolished in strains encoding only the mutated AGA motif, HTkA<sup>G17D</sup>.

We demonstrate here, using comparative differential RNA-seq analyses of strains with unique AHCP landscapes, that



**FIGURE 1** | A single wild-type histone protein is sufficient for normal DNA protection in *T. kodakarensis*. **(A)** Diagrammatic representation of wildtype chromatin modeled from the archaeal histone-based chromatin crystal structure: 9 polymerized histone B dimers (pale green) wrapped by DNA (gray) adapted from Mattioli et al. (2017). The central glycine in the AGA motif at the Loop1–Loop1 interface is colored in red. A Logo-plot highlights the conservation of this motif. Histone dimers may be heterogeneously composed. **(B)** DNA fragments resulting from micrococcal nuclease (MNase) digested chromatin demonstrate the state of chromatin structure in TS600, TS622, and TS620. Chromatin purified from TS600 (TK1413<sup>WT</sup>:histone A/ΔTK2289:histone B) resists MNase digestion over time, resulting in a laddered DNA banding pattern. Prominent 60 and 90 bp bands in addition to higher molecular weight bands (increasing 30 bp increments up to ~300 bp) represent varying levels of histone dimerization and MNase protection. Chromatin purified from TS622 (TK1413<sup>G17D</sup>:histone A/TK2289<sup>WT</sup>:histone B) exhibits an identical protection pattern to TS600 despite encoding a variant (G17D) histone A. This suggests a single WT histone is sufficient for normal chromatin structure formation. Chromatin purified from TS620 (TK1413<sup>G17D</sup>:histone A/ΔTK2289:histone B) exhibits a markedly different protection pattern from TS600 and TS622. The presence of only a variant (G17D) histone A results in a loss of DNAs protected above 90 bp, demonstrating the disruption of the L1–L1 interface interferes with continued histone dimer polymerization. **(C)** Diagrammatic representation of the potential chromatin structures in TS600, TS622, and TS620.

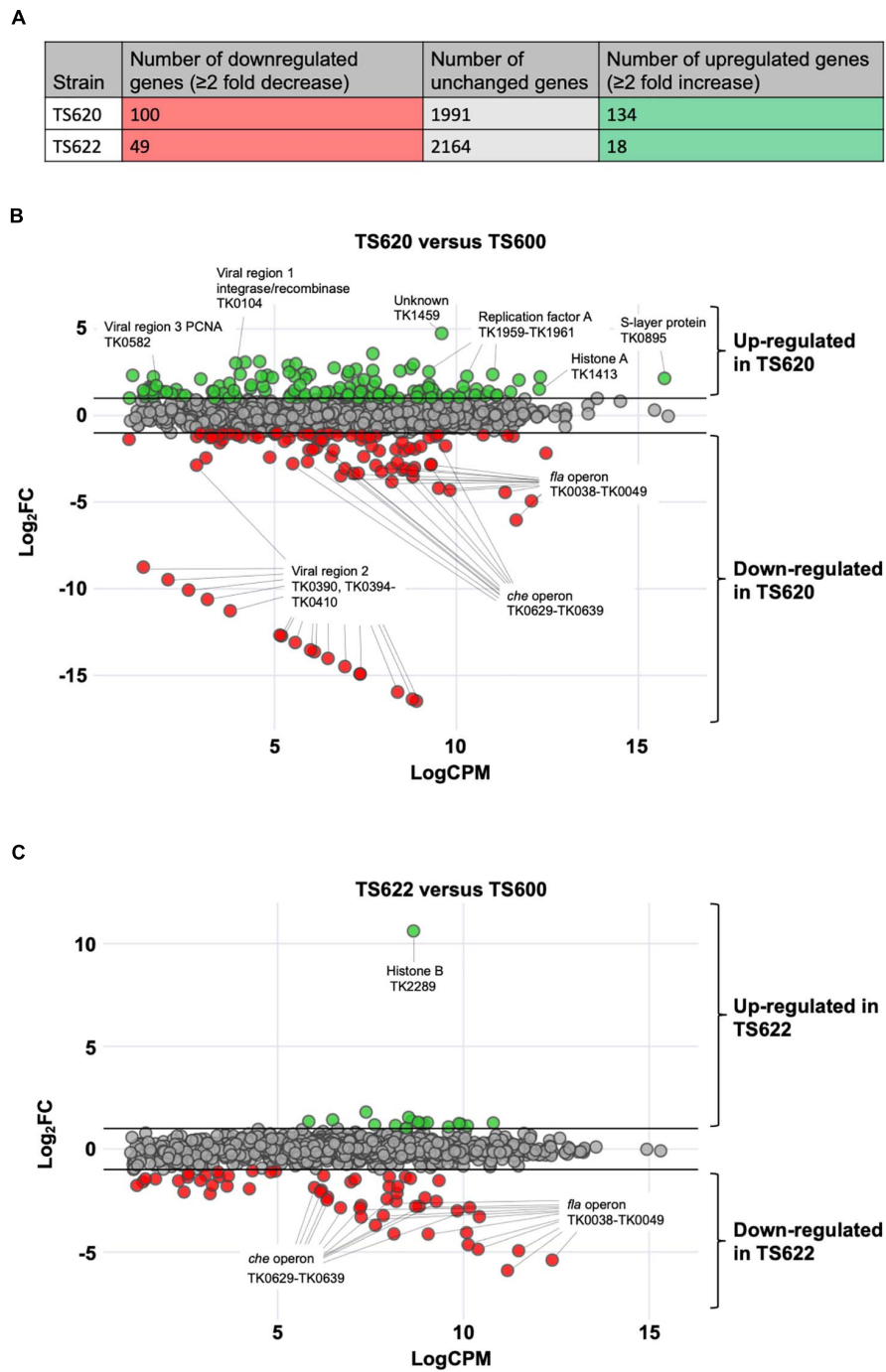
substantial and genome-wide variations in gene expression result from alternating archaeal histone-based chromatin structures, underscoring the importance of AHCPs in normal regulation of gene expression. Expression differences observed in strains lacking extended AHCPs suggest architectural changes in AHCPs are most impactful for the proper regulated expression of chemotaxis-, motility- and proviral-encoding regions of the genome. The global regulatory potential of AHCPs is confirmed and offers impactful routes to control archaeal gene expression by modulating chromatin architectures and AHCP formation *in vivo*.

## MATERIALS AND METHODS

### Strain Construction and Growth Conditions

*Thermococcus kodakarensis* strains were constructed essentially as previously described (Gehring et al., 2017). While *T. kodakarensis* strains deleted for either TK2289 or TK1413 remain viable, strains deleted for TK1413 are no longer

genetically accessible (Čuboňová et al., 2012). Thus, to generate single histone-encoding strains with substitutions that impact AHCP formation, we first deleted TK2289, then introduced allelic changes to TK1413. *T. kodakarensis* strain TS600 was constructed from parental strain TS559 by markerless deletion of TK2289 (HTkB). Strain TS622 was constructed from TS600 by allelic substitution of TK1413 (HTkA) for TK1413<sup>G17D</sup> (HTkA<sup>G17D</sup>). TS620 was constructed by markerless deletion of TK2289 (HTkB) from TS622. Allelic substitution of TK1413 for TK1413<sup>G17D</sup> was confirmed by PCR amplification of the TK1413 loci and subsequent sequencing while deletion of TK2289 was confirmed via diagnostic PCR with primers flanking TK2289. Cultures were grown at 85°C in artificial seawater (ASW) supplemented with 0.5% (w/v) tryptone, 0.5% (w/v) yeast extract (ASW-YT), trace mineral solution and vitamin mixture (nutrient rich medium). ASW contains, per l, 20 g NaCl, 3 g MgCl<sub>2</sub>·6H<sub>2</sub>O, 6 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 1 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 200 mg NaHCO<sub>3</sub>, 300 mg CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.5 g KCl, 420 mg KH<sub>2</sub>PO<sub>4</sub>, 50 mg NaBr, 20 mg SrCl<sub>2</sub>·6H<sub>2</sub>O, and 10 mg Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O. The trace mineral solution (1,000×) contains, per l, 0.5 g MnSO<sub>4</sub>·6H<sub>2</sub>O, 0.1 g CoCl<sub>2</sub>·6H<sub>2</sub>O, 0.1 g ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.01 g



**FIGURE 2 |** Altering 3-dimensional chromatin structure dramatically alters gene expression in *T. kodakarensis*. **(A)** In TS620 (compared to TS600) 100 protein-coding genes were downregulated while 134 protein-coding genes were upregulated. In TS622 (compared to TS600), 49 protein-coding genes were downregulated while 18 protein-coding genes were upregulated. **(B)** In TS620 (TK1413<sup>G17D</sup>:histone A/ $\Delta$ TK2289:histone B) differential RNA-sequencing, represented in an MA plot, revealed a number of genes are significantly upregulated and downregulated when compared to TS600 (TK1413<sup>WT</sup>:histone A/ $\Delta$ TK2289:histone B). Green transcripts are significantly ( $\geq 2$  fold change) enriched in TS620 when compared to TS600. Red transcripts are significantly depleted ( $\geq 2$  fold change) in TS620 when compared to TS600. Notably, in TS620, a number of single stranded binding proteins (TK1959-1961: replication protein A subunits Rpa32, Rpa14, and Rpa41) are significantly enriched in TS620. TK1413<sup>G17D</sup> was found to be enriched in TS620 when compared to TK1413<sup>WT</sup> in TS600. In TS620 transcripts involved in cell motility and cell signaling (TK0038-TK0049: archaeal Fla operon and archaeal che operon (TK0629-TK0639) were significantly depleted. Additionally, a portion (TK0394-TK0410) of viral region 2 (TK0381-TK0421) was significantly depleted. **(C)** In TS622 (TK1413<sup>G17D</sup>:histone A/TK2289<sup>WT</sup>:histone B) far fewer genes were upregulated when compared to TS600. Downregulated genes related in cell motility and environmental signal sensing revealed similar trends to TS620 (*che* and *fla* operons).

**TABLE T1A** | Transcripts enriched and depleted in TS620 (TK1413<sup>G17D</sup>:histone A/ $\Delta$ TK2289:histone B) compared to TS600 (TK1413<sup>WT</sup>:histone A/ $\Delta$ TK2289:histone B). **(A)** The 30 most enriched transcripts in TS620 compared to TS600. **(B)** The 30 most depleted transcripts (*sans* viral region 2 transcripts) in TS620 compared to TS600.

Transcript	Annotation	Process	Fold Change
TK1459	Hypothetical protein	Unknown	22.52
TK1358	Hypothetical protein	Unknown	21.63
TK2061	Sodium/phosphate symporter	Transport	12.84
TK0604	Hypothetical protein	Unknown, viral region 3	12.00
TK0605	Hypothetical protein	Unknown, viral region 3	9.75
TK0208	Phosphoribosyl formylglycinamide cyclo-igase	Purine metabolism	9.51
TK0202	Phosphoribosyl formylglycinamide ynthase, PurS component	Purine metabolism	9.18
TK0204	Phosphoribosylamine-glycine ligase	Purine metabolism	8.97
TK1392	NADH oxidase	Metabolsim	8.74
TK0203	Phosphoribosyl formylglycinamide cyclo-ligase	Purine metabolism	8.43
TK1356	ATPase, AAA superfamily	Unknown, viral region 4	8.41
TK0835	Phosphoribosyl laminoimidazole carboxylase, ATPase subunit	Purine metabolism	7.60
TK1393	Anaerobic glycerol 3-phosphate dehydrogenase	Lipid metabolism	7.40
TK2060	Distant homolog of phosphate transport system regulator PhoU	Transport	7.20
TK1391	Molybdopterin oxidoreductase, 4Fe-4S cluster-binding subunit	Central metabolsim	6.81
TK1023	Hypothetical protein	Unknown	6.64
TK0207	Format-dependent hosporibosylglycinamide formyltransferase	Purine metabolism	6.42
TK1960	Replication factor A complex, RPA14 subunit	Replication/recombination/repair	6.42
TK0201	Phosphoribosyl lformylglycinamide synthase I	Purine metabolism	6.26
TK1464	Hypothetical protein	Unknown	6.05
TK0599	Hypothetical protein	Unknown, viral region 3	5.70
TK1959	Replication factor A complex, RPA32 subunit	Replication/recombination/repair	5.63
TK0389	Hypothetical protein	Unknown, viral region 2	5.57

(Continued)

**TABLE T1A** | Continued

Transcript	Annotation	Process	Fold Change
TK0252	Indole-3-glycerol phosphate synthase	Amino acid synthesis	5.57
TK0836	Phosphoribosyl aminoimidazole carboxylase, catalytic subunit	Purine metabolism	5.50
TK0601	ATPase, AAA superfamily	Unknown, viral region 3	5.42
TK0580	Hypothetical protein	Unknown, viral region 3	5.39
TK0382	Hypothetical protein	Unknown, viral region 2	5.17
TK1961	Replication factor A complex, RPA41 ubunit	Replication/recombination/repair	5.14
TK0600	Hypothetical protein	Unknown, viral region 3	5.13

CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.01 g AlK(SO<sub>4</sub>)<sub>2</sub>·12H<sub>2</sub>O, 0.01 g H<sub>3</sub>BO<sub>3</sub>, and 0.01 g Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O. The vitamin mixture (200×) contains, per l, 0.2 g niacin, 0.08 g biotin, 0.2 g pantothenate, 0.2 g lipoic acid, 0.08 g folic acid, 0.2 g p-aminobenzoic acid, 0.2 g thiamine, 0.2 g riboflavin, 0.2 g pyridoxine, and 0.2 g cobalamin. When present, sodium pyruvate was added at 5 g per l, agmatine sulfate to 1 mM and 6- methyl purine (6MP) to 100 μM (ASW-YT-Pyr). Elemental sulfur (So) was added at 2 g per l in liquid media (ASW-YT-) but was replaced by polysulfides in solid media. Polysulfide solution (500×) contained, per l, 66.7 g sodium sulfide (Na<sub>2</sub>S·9H<sub>2</sub>O) and 3 g sulfur. Gelrite was added to 1% (w/v) to solidify media.

## Chromatin Isolation and Micrococcal Nuclease Digestion

Chromatin isolation and micrococcal nuclease digestions were adapted from Mattioli et al. (2017). Strains TS600, TS620 and TS622 were individual grown to an optical density (measured at 600 nm) of ~0.5 in liquid ASW-YT-S and each starter culture was used to inoculate (1:100) 200 ml of ASW-YT-Pyr cultures that were allowed to grow to an optical density (measured at 600 nm) of ~0.5 to encourage chromatin reprogramming. Cultures were pelleted at ~8,000 × g and immediately frozen at -80°C. Cell pellets were resuspended in 1.0 mL of MNase buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 1 mM CaCl<sub>2</sub>) per 0.2 g of cell mass and ground to homogeneity with a mortar and pestle. Homogenized cells were mechanically lysed by repeated liquid nitrogen freezing and subsequent grinding five times. Whole cell lysate was gently clarified at 1,700 × g for 5 min and the chromatin containing clarified lysate was RNase A digested (Sigma, 4,000 U) for 1 h at 37°C. ~1,500 U of micrococcal nuclease (New England Biolabs) was added to chromatin and aliquots (~100 μl) of digested DNAs were extracted by the addition of 300 μl of 10 mM Tris-HCl pH 8.0 and 400 μl of phenol/chloroform/isoamyl alcohol (25:24:1). Following thorough emulsion and centrifugation at 10,000 × g

TABLE T1B |

Transcript	Annotation	Process	Fold change
TK0038	Archaeal flagellin B1 precursor	Cell motility	-65.39
TK0039	Archaeal flagellin B2 precursor	Cell motility	-30.65
TK0040	Archaeal flagellin B3 precursor	Cell motility	-21.73
TK0042	Archaeal flagellin B5 precursor	Cell motility	-19.94
TK0041	Archaeal flagellin B4 precursor	Cell motility	-18.37
TK0043	Archaeal flagella-related protein C	Cell motility	-14.30
TK0044	Archaeal flagella-related protein D, internal insertion	Cell motility	-11.59
TK0045	Archaeal flagella-related protein F	Cell motility	-11.25
TK0046	Archaeal flagella-related protein G	Cell motility	-10.23
TK0632	Chemotaxis response regulator, CheY	Environmental information processing	-10.09
TK0631	Chemotaxis protein methyltransferase, CheR	Environmental information processing	-9.51
TK0634	Chemotaxis histidine kinase, CheA	Environmental information processing	-9.16
TK0633	Chemotaxis protein-glutamate methyltransferase, containing receiver domain	Environmental information processing	-8.81
TK0635	Chemotaxis histidine kinase	Environmental information processing	-8.39
TK0156	Methyl-accepting chemotaxis protein	Environmental information processing	-8.34
TK2147	Methyl-accepting chemotaxis protein	Environmental information processing	-8.25
TK0049	Archaeal flagella-related membrane protein J	Cell motility	-8.09
TK0047	Archaeal flagella-related protein H	Cell motility	-7.39
TK0630	Methyl-accepting chemotaxis protein	Environmental information processing	-7.14
TK0048	Archaeal flagella-related protein I	Cell motility	-7.13
TK0637	Chemotaxis protein cheC	Environmental information processing	-6.89
TK0050	Hypothetical membrane protein	Unknown	-6.51
TK0636	Chemotaxis protein CheC	Environmental information processing	-6.36

(Continued)

TABLE T1B | Continued

Transcript	Annotation	Process	Fold change
TK0168	Predicted transcription regulator, Lrp/AsnC family	Transcription	-5.47
TK0546	Hypothetical protein	Unknown	-5.32
TK0638	Methyl-accepting chemotaxis protein	Environmental information processing	-5.21
TK1139	ATPase, AAA superfamily	Unknown	-5.19
TK1804	ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	Transport	-4.50
TK1606	Methyl-accepting chemotaxis protein	Environmental information processing	-4.11
TK1605	Hydrolase, metallo-beta-lactamase superfamily	Unknown	-4.09

for 5 min, ~200  $\mu$ L of the DNA-containing aqueous layer was precipitated by the addition of an equal volume of 1 M Tris-HCl pH 8.0 and 2.6 volumes of 100% ethanol proceeding a 1-h incubation at  $-80^{\circ}\text{C}$ . DNAs were pelleted in a  $4^{\circ}\text{C}$  centrifuge at  $10,000 \times g$  for 30 min and subsequently resolved in a 4% agarose gel.

## RNA Isolation

TS600 and TS622 were grown in triplicate to an optical density (measured at 600 nm) of ~0.5 in liquid ASW-YT-S and used to inoculate (1:100) 300 ml of ASW-YT-Pyr per strain and allowed to grow to an optical density (measured at 600 nm) of ~0.05 to encourage chromatin reprogramming. Cultures were rapidly chilled and pelleted at  $8,000 \times g$  for 5 min and then resuspended in 1.0 mL of Trizol (Invitrogen) with a 10-min incubation at room temperature. A total of 200  $\mu$ l of chloroform was added followed by centrifugation at  $10,000 \times g$  at  $4^{\circ}\text{C}$  for 15 min yielding an RNA-containing aqueous layer which was added to 500  $\mu$ l of isopropanol and incubated at room temperature for 10 min. Centrifugation at  $10,000 \times g$  for 15 min at  $4^{\circ}\text{C}$  produced an RNA pellet that was washed with 1 ml of 75% ethanol and subsequently resuspended in 88  $\mu$ l of RNase-free  $\text{H}_2\text{O}$ , 10  $\mu$ l of DNaseI buffer, and 1  $\mu$ l of DNaseI (New England Biolabs) to digest residual DNA ( $37^{\circ}\text{C}$  for 30 min). Replicate samples were prepared identically.

## RNA-Seq Library Preparation

### TS600 and TS622

Using ~120 ng of RNA that was depleted of rRNA, following the NEBNext<sup>®</sup> rRNA Depletion Kit (E6310), cDNA libraries were constructed with New England Biolabs NEBNext<sup>®</sup> Ultra<sup>™</sup> Directional RNA Library Prep kit for Illumina<sup>®</sup> (E7420s) and NEBNext<sup>®</sup> Multiplex Oligos for Illumina<sup>®</sup> (Index primer set 1, E7335) according to the manufacturer's procedure. The multiplexed libraries were sequenced at Cofactor Genomics using

**TABLE T2A** | Transcripts enriched and depleted in TS622 (TK1413<sup>G17D</sup>:histone A/TK2289<sup>WT</sup>:histone B) compared to TS600 (TK1413<sup>WT</sup>:histone A/ $\Delta$ TK2289:histone B). **(A)** All transcripts enriched in TS622 compared to TS600. Histone B, TK2289 denoted by \* is absent in TS600 resulting in a large fold change. **(B)** The 30 most depleted transcripts in TS622 compared to TS600.

<b>A</b>			
Transcript	Annotation	Process	Fold change
TK2289	Archaeal histone B	Chromatin	1565.20
TK1020	Hypothetical membrane protein	Unknown	3.49
TK0717	Molybdate transport system substrate-binding protein	Transport	2.93
TK0162	Hypothetical membrane protein	Unknown	2.69
TK0718	Molybdate transport system permease protein	Transport	2.58
TK0720	Hypothetical protein	Unknown	2.55
TK0166	Hypothetical protein	Unknown	2.48
TK2070	Sulfhydrogenase subunit delta	Energy Uetabolism	2.45
TK1862	Hypothetical protein	Unknown	2.45
TK0467	Hypothetical protein	Unknown	2.42
TK2071	Sulfhydrogenase subunit gamma (sulfur reductase)	Energy metabolism	2.37
TK0719	Molybdate transport system ATP-binding protein	Transport	2.37
TK2072	Sulfhydrogenase subunit beta (sulfur reductase)	Energy metabolism	2.33
TK0164	S-layer-like array protein	Cell structure	2.33
TK0163	ABC-2 type transport system permease protein	Transport	2.27
TK1463	Hypothetical protein	Unknown	2.22
TK2278	Myo-inositol-1-phosphate synthase	Metabolism	2.19
TK0765	Glyceraldehyde-3-phosphate dehydrogenase (NAD(P))	Metabolism	2.12
TK2069	Sulfhydrogenase subunit alpha	Energy metabolism	2.04

one high output NextSeq Illumina® run for single-end reads with a minimum read length of 75 bp and with a requested >40 million reads per sample.

### TS600 and TS620

A total of 1.5  $\mu$ g of RNA was processed at Novogene for Prokaryotic RNA-seq, specifically for rRNA depletion (RiboZero™ Magnetic Kit), library construction (NEBNext® Ultra™ RNA Library Prep kit) and 150 bp paired-end sequencing on a HiSeq Illumina® platform.

**TABLE T2B** |

<b>B</b>			
Transcript	Annotation	Process	Fold change
TK0038	Archaeal flagellin B1 precursor	Cell motility	-59.54
TK0039	Archaeal flagellin B2 precursor	Cell motility	-41.95
TK0040	Archaeal flagellin B3 precursor	Cell motility	-30.60
TK0042	Archaeal flagellin B5 precursor	Cell motility	-29.13
TK0041	Archaeal flagellin B4 precursor	Cell motility	-25.02
TK0043	Archaeal flagella-related protein C	Cell motility	-17.47
TK0812	Adenylate kinase	Punne metabolism	-17.38
TK0811	Hypothetical protein	Unknown	-16.68
TK0631	Chemotaxis protein methyltransferase CheR	Environmental information processing	-13.03
TK0046	Archaeal flagella-related protein G	Cell motility	-9.90
TK0044	Archaeal flagella-related protein D, internal insertion	Cell motility	-9.76
TK0045	Archaeal flagella-related protein F	Cell motility	-9.40
TK0156	Methyl-accepting chemotaxis protein	Environmental information processing	-8.01
TK0635	Chemotaxis histidine kinase	Environmental information processing	-7.36
TK0632	Chemotaxis protein CheY	Environmental information processing	-7.27
TK0630	Methyl-accepting chemotaxis protein	Environmental information processing	-7.20
TK0048	Archaeal flagella-related protein I, predicted Csecretion ATPase	Cell motility	-6.92
TK0633	Chemotaxis protein-glutamate methyltransferase, containing receiver domain	Environmental information processing	-6.89
TK0047	Archaeal flagella-related protein H, predicted ATPase	Cell motility	-6.71
TK0049	Archaeal flagella-related membrane protein J	Cell motility	-5.85
TK0634	Sensor kinase CheA	Environmental information processing	-5.83
TK0636	Chemotaxis protein CheC	Environmental information processing	-5.59
TK0050	Hypothetical protein	Unknown	-5.39

(Continued)

TABLE T2B | Continued

Transcript	Annotation	Process	Fold change
TK2147	Methyl-accepting chemotaxis protein	Environmental information processing	-5.19
TK0637	Chemotaxis protein CheC	Environmental information processing	-5.07
TK0431	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase	Purine metabolism	-4.51
TK0638	Methyl-accepting chemotaxis protein	Environmental information processing	-4.43
TK0432	Phosphonobosylaminoimidazole-succinocarboxamide synthase	Purine metabolism	-4.26
TK1139	ATPase, AAA superfamily	unknown	-4.23
TK0051	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	unknown	-4.01

## Data Preprocessing TS600 and TS622

RNA-Seq reads were first analyzed for quality control using FastQC (Andrews, 2010). To remove adapter sequences and other artifacts, fastx-trimmer was used to trim the first 11 positions in each read (Gordon). After filtering, the reads were aligned to the *T. kodakarensis* (KOD1) reference genome using bowtie (Langmead, 2010) with the parameter *-m 1*, thus ensuring suppression of all multiply aligned reads. Finally, the bowtie output was converted to BAM format, sorted, and indexed using samtools (Li et al., 2009).

## TS600 and TS620

RNA-Seq reads were first analyzed for quality control using FastQC (Andrews, 2010). To remove adapter sequences and other artifacts, fastx-trimmer was used to trim the first 19 positions in each read (Gordon). After filtering, the reads were aligned to the *T. kodakarensis* (KOD1) reference genome using bowtie2 (Langmead and Salzberg, 2012) with default parameters. Next, in each library, reads that aligned to multiple locations were filtered out. Finally, the filtered output was converted to BAM format, sorted, and indexed using samtools (Li et al., 2009).

## Differential Gene Expression Analysis

To identify differentially expressed genes we used EdgeR (Robinson et al., 2009). Read counts for every gene in the Ensembl annotations of the species were generated using a custom python script that used SpliceGrapher (Rogers et al., 2012) and pysam (Heger et al., 2014). The EdgeR *p*-values were adjusted for multiple comparisons using the Benjamini-Hochberg method

(Benjamini and Hochberg, 1995). Finally, we used a cutoff of 1.00 on the transcript abundance (log CPM) and fold change (log FC).

## Data Visualization and Plotting

To visualize differential gene expression, we used Plotly (Sievert et al., 2017) to generate the MA plots. All analyses are provided in the supplementary jupyter notebooks. Circos (Krzywinski et al., 2009) plots were generated using the FPKM values of genes, measured using stringtie2 tool (Pertea et al., 2015) using the Ensembl gene annotations.

## DNA Sequencing of TS620

Genomic DNA was purified from strain TS620 using the Monarch Genomic DNA Purification kit (New England Biolabs). Pacific Biosciences (PacBio) libraries were constructed following the Pacific Biosciences Template Preparation and Sequencing Protocol. The library was then sequenced on a PacBio Sequel instrument using Polymerase 3.0 Chemistry and diffusion loading for 600 min. The data was analyzed using PacBio SMRT Analysis tools.

## RESULTS

### A Single Histone Protein Is Sufficient for AHCP Formation

To assess the role of chromatin superstructure, reflected by the totality of AHCPs, on gene expression in living cells, we generated *T. kodakarensis* strains that encode histone variants known to impact AHCP length. While strains retaining only a single histone isoform are viable, reducing the concentration of total histone proteins (Čuboňová et al., 2012; Mattioli et al., 2017; Sanders et al., 2019b), or retaining histone variants with dramatically reduced DNA binding affinity (Čuboňová et al., 2012; Mattioli et al., 2017) is not tolerated in *T. kodakarensis*. We thus carefully constructed strains wherein we retained the natural expression profile of one or both histone proteins but altered the sequence of individual histone-encoding genes to produce proteins at native levels (Mattioli et al., 2017) that impact AHCP formation. Strains encoding only a single histone isoform in WT or variant form (Supplementary Figure S1) were generated, as were strains wherein one isoform remained WT, and the other was modified to replace G17 with D. Markerless modification or deletion of the genes encoding the endogenous histone proteins, HTkA (TK1413) and HTkB (TK2289), in their natural context allowed for preservation of the native promoter elements and did not impact the surrounding loci. Strain TS559 served as the parental strain for construction of three strains with varied histone compositions. Deletion of TK2289 (encoding HTkB) resulted in strain TS600 (encoding only HTkA<sup>WT</sup>). Modification of TK1413 to generate a G17D variant of HTkA in an otherwise native background, including the presence of WT HTkB, generated strain TS622. TS620 combines both genetic modifications, thereby generating a strain supported by only the synthesis of a single, G17D HTkA variant. Deletion of TK2289, encoding HTkB, was confirmed in TS600



and TS620 by PCR amplification of the flanking genomic regions and sequencing to confirm the exact desired endpoints of the genomic modification (**Supplementary Figure S1B**). HTkA<sup>G17D</sup> variants were confirmed in TS620 and TS622 by amplifying the entire HTkA coding and promoter sequences followed by Sanger sequencing to confirm retention of all native regulatory and coding sequences with the exception of the desired missense mutation (**Supplementary Figures S1B,C**).

To assess the impacts of varied histone isoforms on AHCP formation, total chromatin purified from strains TS600, TS620, and TS622 was subjected to micrococcal nuclease (MNase) digestion. MNase digestion provides a rapid, genome-wide mechanism to define minimal units of chromatin - such as the nucleosome in Eukarya. Digestion of chromatin isolated from TS600 confirmed the previously observed wildtype AHCP protection patterning - a prominent ~60 bp DNA fragment along with larger DNA fragments in increasing ~30 bp increments (up to ~300 bp) - indicative of varied-length AHCPs (Čuboňová et al., 2012; Mattioli et al., 2017; Henneman et al., 2018; Rojec et al., 2019; Henneman et al., 2020; Stevens et al., 2020) (**Figures 1B,C**). The persistence of the distinct 60 + 30(n) ladder demonstrates that AHCPs are varied in length but generally stable, thus providing architectures that can be exploited to regulate DNA accessibility and gene expression. TS600 encodes only the HTkA isoform, therefore all histone-based chromatin structure in this strain is composed of this single histone (**Figure 1C**).

In contrast, when HTkA was modified to place a larger and charged residue within the AGA motif of L1 (G17D), the chromatin from strain TS620 displayed a dramatically different MNase protection pattern (**Figure 1B**). Discrete DNA fragments >90 bp were absent, demonstrating histone:DNA interactions occurred to allow initial DNA wrapping but that continued polymerization to form extended AHCPs was not possible. The observed digestion pattern is consistent with previous (Mattioli et al., 2017) digestions of chromatin from variant HTkA<sup>G17D</sup>-encoding strains, suggesting histone dimers form tetramers, protecting ~60 bp of DNA, and an additional dimer interacts to form a hexamer, protecting ~90 bp of DNA, but that larger associations of histone dimers are restricted due to clashes between adjacent gyres of AHCPs and loss of potential electrostatic interactions across the adjacent gyres (Henneman et al., 2018; Henneman et al., 2020; Bowerman et al., 2021). Thus, across the entire genome, the single HTkA<sup>G17D</sup> variant encoded in TS620 disrupts the L1-L1 interface within AHCPs, preventing continued polymerization of histone dimers that normally provides a route to extended AHCP formation.

Surprisingly, digestion of total chromatin purified from TS622 resulted in an MNase protection pattern nearly identical to TS600 (**Figure 1B**) suggesting the presence of variant HTkA<sup>G17D</sup> did not interfere with superhelix formation of the HTkB isoform. Given that chromatin from strain TS620 demonstrates HTkA<sup>G17D</sup> alone cannot make up the superhelix, it is likely that all the larger regions of DNA protection in strain TS622 result from superstructures formed entirely of HTkB. It is possible that at least some of the smaller (60–90 bp) protected DNA fragments result from chromatin structures composed entirely of

HTkA<sup>G17D</sup> or from HTkB/HTkA<sup>G17D</sup> heteromers; additionally, it is possible HTkA<sup>G17D</sup> may cap or terminate the superhelices composed of HTkB resulting in protection of >90 bp fragments (Henneman et al., 2020; Stevens et al., 2020) (**Figures 1B,C**).

## AHCP Structures Regulate Genome-Wide Gene Expression

With histone composition altered in strain TS622, and histone-composition and extended AHCP formation disrupted in strain TS620, we sought to quantify the transcriptomes of each strain in response to an environmental shift (Jäger et al., 2014; Atomi and Reeve, 2019). Environmentally cued changes to histone-based chromatin architecture are a known mechanism to regulate gene expression, and for the *Thermococcales*, one of the largest determinants of metabolism and gene expression profiles is the availability of different terminal electron acceptors (Jäger et al., 2014; Mattioli et al., 2017). Strains with variant histone- and AHCP-landscapes were grown to early exponential phase under conditions that permit elemental sulfur (S<sub>0</sub>) to serve as the terminal-electron acceptor, then rapidly transferred to conditions wherein S<sub>0</sub> was absent, signaling a metabolic shift requiring substantial changes in gene expression for continued rapid growth.

RNAs from each strain were purified from cultures following the environmental shift, depleted of ribosomal RNAs, and subjected to RNA-sequencing to quantify steady-state transcript abundance. RNAs were collected at early time points following the environmental shift to monitor the primary impacts of variant AHCP architectures on gene expression. Differential expression analyses defined meaningful changes in the transcriptomes of TS620 and TS622 compared to TS600 (**Figure 2**). Comparison of transcriptomes resultant from TS600 (HTkA<sup>WT</sup> only; typical AHCPs of varied lengths) and TS620 (HTkA<sup>G17D</sup> only; containing only short AHCPs) provides a quantitative measure of the regulation normally afforded by AHCPs in *T. kodakarensis* (**Figures 1C, 2**). In contrast, comparison of the transcriptomes of TS600 and TS622 (HTkB<sup>WT</sup>, HTkA<sup>G17D</sup>) highlights the impact of histone variants on AHCPs that can regularly form with a mixture of HTkB and HTkA<sup>G17D</sup> (**Figures 1C, 2**). Transcriptome changes were quantified by comparing log<sub>2</sub>-average fold change (Log<sub>2</sub>FC) against log-average counts per million (LogCPM) (**Figures 2B,C**). The *T. kodakarensis* genome encodes ~2,300 annotated open reading frames, with abundant antisense transcription and many small transcripts (Fukui et al., 2005; Jäger et al., 2014; Sas-Chen et al., 2019). Our sequencing coverage was sufficient to detect nearly all transcripts with a LogCPM > 1.00: 2,225 transcripts were included in the TS620/TS600 comparison; 2,231 transcripts were included in the TS622/TS600 comparison.

The substitution of HTkA<sup>G17D</sup> (TS620) for HTkA<sup>WT</sup> (TS600) as the only histone in *T. kodakarensis* resulted in significant (> 2-fold) changes in the steady-state abundance of 234 genes, representing ~11% of the entire transcriptome. The lack of extended AHCPs in TS620 resulted in increased abundance of 100 genes and decreased abundance of another 134 genes (**Figure 2A**); genes with both ordinarily high or

low expression were differentially expressed (Jäger et al., 2014). Genes-encoding proteins involved in central metabolism, purine synthesis and metabolism, amino acid synthesis, transport and a number of hypothetical proteins were upregulated (Table 1A). No obvious chromosomal distribution was noted, with the obvious exception of coregulation of operons, suggesting the entire chromosome is normally subject to regulation imposed by extended AHCPs. Further, hinting that expression of histone-encoding genes are regulated by AHCPs, TK1413<sup>G17D</sup> transcripts were enriched ~2.8 fold in TS620 compared to TK1413<sup>WT</sup> in TS600. Altered AHCP structures undoubtedly impact DNA replication, recombination and repair, and the loss of extended AHCPs results in the large increase in abundance of all three replication factor A proteins (archaeal RPA is a heterotrimer composed in *T. kodakarensis* of the products of TK1959 (increased ~5.2 fold), TK1960 (increased ~5.8 fold) and TK1961 (increased ~4.8 fold) in strain TS620. The likely increased abundance of functional RPA proteins suggests inhibiting extended AHCPs may permit some DNA regions to locally unwind or melt, thereby requiring more RPA to protect the increased abundance of single-stranded DNA (Figure 2B).

While many gene classes showed increases in transcript abundance due to changes in AHCP structure, a large percentage of genes related to cell motility and environmental signal sensing were downregulated in the absence of extended AHCPs in TS620 when compared to TS600. Localized AHCPs are likely to regulate gene expression both positively and negatively, depending on the availability of DNA sequences critical for gene expression. Expression of operons encoding archaeal components (annotated as the *fla* operon) and chemotaxis proteins (*che* operon) appear particularly sensitive to AHCPs and are likely dependent on AHCPs for proper regulation (Figure 2B). Genes comprising the entire *T. kodakarensis* archaeal-encoding *fla* operon (TK0038-TK0049) and chemotaxis-encoding *che* operon (TK0629-TK0639) were ~7.1–65-fold and ~2.7–10-fold less abundant in strain TS620 compared to strain TS600 (Table 1B).

Given that AHCPs can be generated in strain TS622 – as assessed by MNase digestions (Figure 1) – we predicted a more minor impact on the total transcriptome of strain TS622 compared to TS600. Despite the presence of HTkA<sup>G17D</sup>, the added presence of HTkB in TS622 permits sufficient AHCP formation to reduce the number of aberrantly transcribed genes when compared to TS620. When comparing the transcriptomes of TS622 (HTkA<sup>G17D</sup>, HTkB) and TS600 (HTkA<sup>WT</sup>, ΔHTkB), we noted only approximately half as many transcripts (49) of decreased abundance in TS622, while just 18 transcripts were enriched (Figure 2C, Table 2A). Similar decreases in the abundance of transcripts encoding cell motility and environmental signal sensing were observed in TS622, with transcripts from the *fla*- and *che*-operon decreased ~6–59-fold and ~2.6–9-fold, respectively (Figure 2C, Table 2B). The decreased abundance of *fla*- and *che*-operon transcripts suggests the HTkA isoform is critical for proper regulation of these loci or factors that control expression of such loci and that the presence of HTkB<sup>WT</sup> cannot compensate for the loss of HTkA<sup>WT</sup> in these limited scenarios.

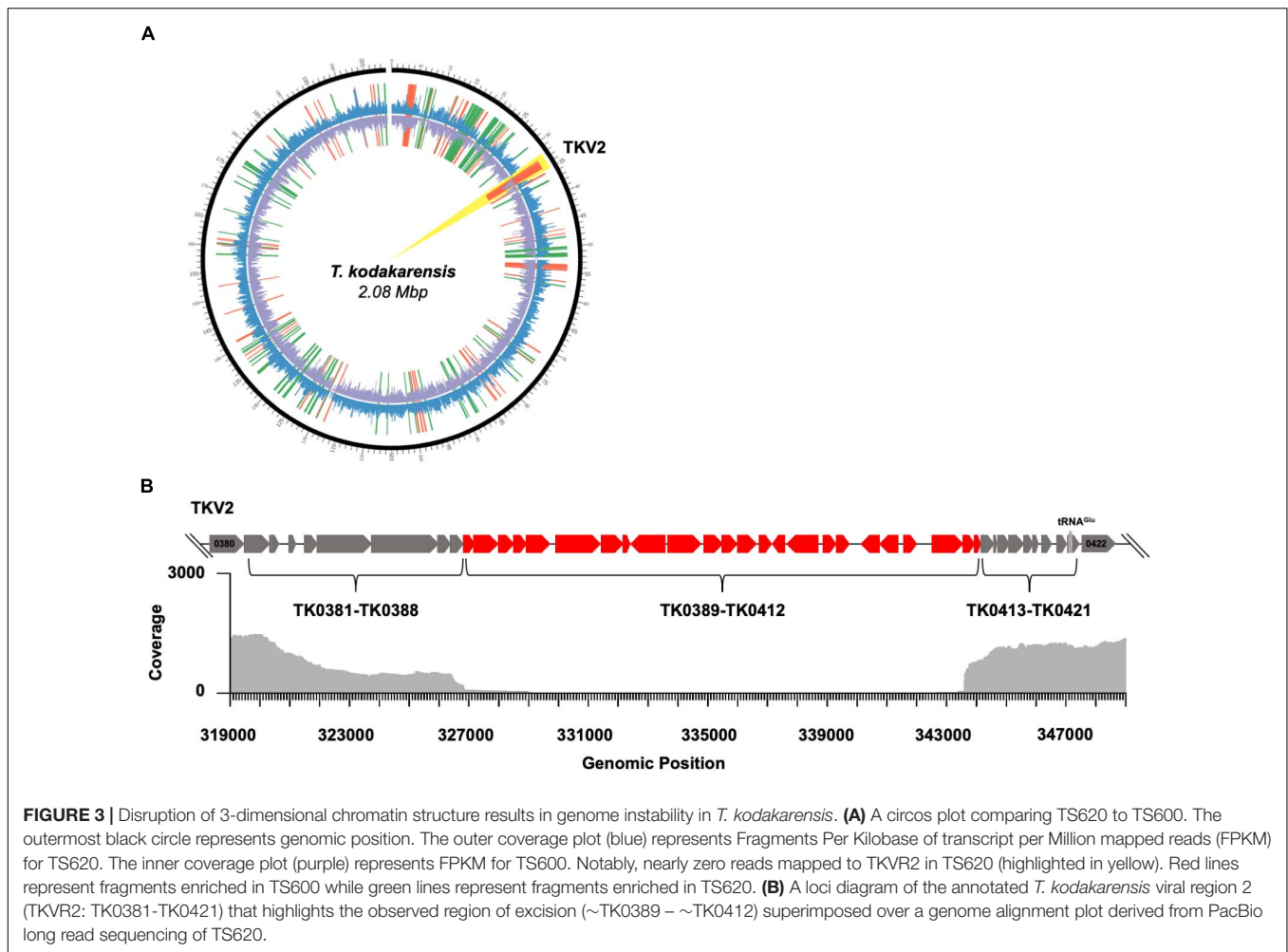
## AHCPs Are Necessary for Proviral Region Expression and Retention

The *T. kodakarensis* genome contains 4 annotated, non-essential proviral regions, each ~20–25 Kbp in length: TKV1 (TK0073-TK0105), TKV2 (TK0381-TK0421), TKV3 (TK0575-TK0614), and TKV4 (TK1342-TK1378) (Fukui et al., 2005; Tagashira et al., 2013). Comparisons of the genomes of many *Thermococcales* suggests recombination events between proviral regions can rearrange the genome context in evolutionary timescales; however, the proviral regions of the *T. kodakarensis* appear genetically stable (Krupovic et al., 2013; Cossu et al., 2015, 2017). These proviral regions represent ~100 Kbp in total (~5% of the 2.08 Mbp genome) and encode a plethora of genes with unknown function. Expression of the proviral regions has been observed in all transcriptome studies of *T. kodakarensis* (Tagashira et al., 2013; Jäger et al., 2014). An observed trend in the transcriptomic profiles obtained from strain TS620 was the depletion or enrichment of genes assigned to the *T. kodakarensis* viral regions (Figures 2B,D). Several predicted integrase genes (TK0104 and TK0381) were upregulated in TS620 as well as the non-essential PCNA2 (Pan et al., 2013) (TK0582) and a predicted AAA superfamily ATPase (TK0601), (Supplementary Figures 3A,B). Most intriguing, transcripts aligning to a central portion of TKV2 (TK0390, TK0394-TK0410) were completely depleted from TS620. Among these genes were several predicted SpoVT, AbrB transcriptional regulators, and many hypothetical genes (TK0402, TK0405, TK0406, TK0409, and from TKV4, TK1372).

The observed complete absence of these TKV2 transcripts prompted further evaluation of the genome of TS620 (Figure 3). Whole-genome sequencing (WGS) of TS620 revealed a relatively large (~15 Kbp) central region of TKV2 was spontaneously excised from the genome (Figure 3B). Excision of most, but not all of TKV2 was confirmed by PCR amplifications of loci within and flanking TKV2 sequences in the genome of TS620 (Supplementary Figure S2). This missing region of TKV2 within strain TS620 genomic sequences aligns closely with the observed lack of reads aligning to TKV2 (Figure 3A). Despite differential expression of portions of other viral regions, the genomic loci for TKVR1, TKVR3, and TKVR4 remain intact (Supplementary Figure S3). The excision of TKV2 in only TS620 suggests not only are AHCP necessary for regulated gene expression, but that these structures also play a role in genome stability and recombination, perhaps related to viral region retention or repression.

## DISCUSSION

Histone proteins encoded in most Archaea are the primary proteins responsible for genome organization (Nalabothula et al., 2013; Peeters et al., 2015; Henneman et al., 2018; Henneman et al., 2020; Sanders et al., 2019b; Stevens et al., 2020). Despite geometric and structural similarities, the potential for archaeal histones to form a continuous helical polymer is distinct from the eukaryotic nucleosome (Mattiroli et al., 2017). Formation of AHCPs is a major regulatory event in adaptive gene expression in *T. kodakarensis* and the regulation afforded by AHCP structures likely extends to most histone-encoding Archaea



(Bhattacharyya et al., 2018; Sanders et al., 2019b; Henneman et al., 2020; Stevens et al., 2020; Bowerman et al., 2021; Laursen et al., 2021).

The transcriptomes of *T. kodakarensis* strains that can (TS600 and TS622) and cannot generate extended AHCPs (TS620) reveal significant differences most likely associated directly with altered chromatin landscapes. The impacts are often dramatic, and the regulation is not associated with isolated chromosomal regions but instead is noted genome wide. In strain TS620, where formation of stable and detectable (via MNase digestions) AHCPs was inhibited, >11% of the transcriptome was significantly altered. Approximately equal numbers of transcripts were more or less abundant due to the loss of AHCPs, suggesting that AHCPs can both positively and negatively impact gene expression, depending on the loci. Expression of the *che-* and *fla-*operons, as well as many proviral regions were the most dramatically affected by changes in the AHCP formation. In support of AHCP length influencing gene expression in these operons/proviral regions, expression profiles of *T. kodakarensis* strains encoding WT versions of a single histone isoform displayed no significant differences in the expression of these same operons (Čuboňová et al., 2012).

AHCPs composed entirely of HTkA or HTkB are known to be sufficient for normal expression suggesting the presence of HTkA<sup>G17D</sup> disrupts AHCP formation, thereby altering transcription in both TS620 and TS622. It is also plausible disrupted chromatin structures limit or increase expression of transcription factors that regulated select operon expression, but our transcriptomics data does not identify any obvious candidates.

Perhaps the most striking dysregulation observed in strains incapable of chromatin superhelix formation (TS620) was the loss of a portion of TKV2 from the genome. Like many proviral integrations in archaeal genomes, all four proviral regions in the *T. kodakarensis* genome overlap with tRNA encoding loci and large rearrangements noted in the genomes of *T. kodakarensis* and related *Thermococcales* often begin and end internal to the proviral regions (Geslin et al., 2007; Ortmann et al., 2008; Soler et al., 2008; Gonnet et al., 2011; Mochizuki et al., 2011; Gorlas et al., 2012; Tagashira et al., 2013). The combined impacts of disrupted AHCP formation on replication and recombination, the increased abundance of predicted viral integrase transcripts, and the potential for more ssDNA due to DNA melting in strains with altered chromatin

landscapes provides a plausible explanation for the loss of TKV2 sequences. The inability to detect DNA sequences encompassing TK0389-0412 or viral particles containing these genes suggests this DNA fragment was degraded following excision from the genome (Soler et al., 2008). Although dispensable, the obvious growth defects of *T. kodakarensis* strains lacking proviral regions suggest their incorporation and proper regulation within archaeal genomes confers an evolutionary advantage (Hawkins et al., 2013; Tagashira et al., 2013).

The retention of histone proteins in most archaeal clades suggest histone-based chromatin structures provide beneficial regulatory roles that are exploited to provide a level of regulation on gene expression. The varied lengths of AHCPs, the presence of multiple histone isoforms in many Archaea, and the known changes in histone isoform expression in response to environmental changes all present routes to activate, repress and fine-tune gene expression to maximize growth in changing environments (Mattioli et al., 2017; Bhattacharyya et al., 2018; Henneman et al., 2018; Henneman et al., 2020; Sanders et al., 2019b; Stevens et al., 2020). The potential for biologically significant post-translational modification of archaeal histones may provide an additional regulatory avenue (Alpha-Bazin et al., 2020).

The mechanisms controlling the formation of AHCPs with varying lengths at different loci and in different sequence contexts likely plays an additional role in transcription regulation (Henneman et al., 2018; Henneman et al., 2020; Sanders et al., 2019a,b, 2020; Stevens et al., 2020) and may be exploited by yet to be discovered archaeal chromatin remodeling complexes or histone isoforms that promote or inhibit formation of archaeal chromatin superstructures.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study have been deposited in the NIH GEO repository at <https://www.ncbi.nlm.nih.gov> and can be accessed with accession number GSE151920.

## AUTHOR CONTRIBUTIONS

TSd, SF, RV, BB, and TSt designed, constructed, and confirmed the local genotypes of *T. kodakarensis* strains used in this work. TSd purified proteins, DNAs and RNAs, performed MNase digestions, generated RNA-seq libraries, analyzed data,

built structural models, prepared figures, and helped write the manuscript. FU and AB-H analyzed RNA-sequencing datasets, prepared figures and helped write the manuscript. AGE and AGa prepared and sequenced DNA sequencing libraries, analyzed whole genome sequencing results, prepared figures, and helped write the manuscript. TSd and TSt prepared the initial manuscript that was edited, improved, and ultimately approved by all authors. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Institute of Health grant No. GM100329 (To TSt) and by funding from New England Biolabs, Inc. (To AGE and AGa).

## ACKNOWLEDGMENTS

We thank members of the Santangelo, Ben-Hur, and Gardner labs for critical review of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.681150/full#supplementary-material>

**Supplementary Figure 1** | Construction of TS600, TS620, and TS622. **(A)** Genomic loci diagrams for histone A (TK1413) and histone B (TK2289) and the modifications to each in strains TS600, TS620, and TS622 compared to parental strain TS559. **(B)** Amplification of TK1413 and TK2289 using the primers diagramed in **(A)**. The open reading frame for TK2289 has been markerlessly deleted in both TS600 and TS620. Amplicons for TK1413 (lanes 1, 2, and 3) were sequenced and **(C)** confirmed to contain either a glycine codon at position 17 (TS600) or aspartic acid (TS620 and TS622).

**Supplementary Figure 2** | Viral region expression dysregulation results in the partial loss of TKV2. **(A)** All viral region transcripts enriched in TS620 when compared to TS600. **(B)** All viral region transcripts depleted in TS620 when compared to TS600. **(C)** Primer Pairs A/B, C/D, and E/F were used to amplify the upstream, downstream and central region of TKV2 in 1-TS600, 2-TS622, and 3-TS620.

**Supplementary Figure 3** | TKV1, TKV3, and TKV4 remain stably integrated in the *T. kodakarensis* genome. TKV1 spans TK0073-TK0105. TKV3 spans TK0575-TK0614. TKV4 spans TK1342-TK1378.

## REFERENCES

- Alpha-Bazin, B., Gorlas, A., Lagorce, A., Joulie, D., Boyer, J. B., Dutertre, M., et al. (2020). 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
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Atomi, H., and Reeve, J. (2019). Microbe profile: *Thermococcus kodakarensis*: The modehyperthermophilic archaeon. *Microbiol. (U. K.)* 165, 1166–1168. doi: 10.1099/mic.0.000839
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bhattacharyya, S., Mattioli, F., and Luger, K. (2018). Archaeal DNA on the histone merry-go-round. *FEBS J.* 285, 3168–3174. doi: 10.1111/febs.14495

- Bowerman, S., Wereszczynski, J., and Luger, K. (2021). Archaeal chromatin “slinkies” are inherently dynamic complexes with deflected DNA wrapping pathways. *Elife* 10:65587. doi: 10.7554/eLife.65587
- Cossu, M., Badel, C., Catchpole, R., Gabelle, D., Marguet, E., Barbe, V., et al. (2017). Flipping chromosomes in deep-sea archaea. *PLoS Genet.* 13:e1006847. doi: 10.1371/journal.pgen.1006847
- Cossu, M., Da Cunha, V., Toffano-Nioche, C., Forterre, P., and Oberto, J. (2015). Comparative genomics reveals conserved positioning of essential genomic clusters in highly rearranged *Thermococcales* chromosomes. *Biochimie* 118, 313–321. doi: 10.1016/j.biochi.2015.07.008
- Čuboňová, L., Katano, M., Kanai, T., Atomi, H., Reeve, J. N., and Santangelo, T. J. (2012). An archaeal histone is required for transformation of *Thermococcus kodakarensis*. *J. Bacteriol.* 194, 6864–6874. doi: 10.1128/JB.01523-12
- Decanniere, K., Babu, A. M., Sandman, K., Reeve, J. N., and Heinemann, U. (2000). Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermobacter fervidus*. *J. Mol. Biol.* 303, 35–47. doi: 10.1006/jmbi.2000.4104
- Dulmage, K. A., Todor, H., and Schmid, A. K. (2015). Growth-Phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. *MBio* 6:e00649-15. doi: 10.1128/mBio.00649-15
- Farkas, J. A., Picking, J. W., and Santangelo, T. J. (2013). Genetic techniques for the Archaea. *Annu. Rev. Genet.* 47, 539–561. doi: 10.1146/annurev-genet-111212-132225
- Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., and Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* 15, 352–363. doi: 10.1101/gr.3003105
- Gehring, A., Sanders, T., and Santangelo, T. J. (2017). Markerless gene editing in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Bio Prot.* 7:2604. doi: 10.21769/bioprotoc.2604
- Geslin, C., Gaillard, M., Flament, D., Rouault, K., Le Romancer, M., Prieur, D., et al. (2007). Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi*. *J. Bacteriol.* 189, 4510–4519. doi: 10.1128/JB.01896-06
- Gonnet, M., Erauso, G., Prieur, D., and Le Romancer, M. (2011). PAMT11, a novel plasmid isolated from a *Thermococcus* sp. strain closely related to the virus-like integrated element TKV1 of the *Thermococcus kodakaraensis* genome. *Res. Microbiol.* 162, 132–143. doi: 10.1016/j.resmic.2010.11.003
- Gordon, A. Unpublished. G. H.-F. Short-Reads Preprocessing Tools, and 2010, Undefined Fastx-Toolkit. Available online at: [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)
- Gorlas, A., Koonin, E. V., Bienvenu, N., Prieur, D., and Geslin, C. (2012). TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. *Environ. Microbiol.* 14, 503–516. doi: 10.1111/j.1462-2920.2011.02662.x
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature* 503, 544–547. doi: 10.1038/nature12650
- Heger, A., Belgrad, T. G., Goodson, M., and Jacobs, K. (2014). *pysam: Python interface for the SAM/BAM sequence alignment and mapping format*. Available online at: <https://github.com/pysam-developers/pysam> (accessed February 1, 2016).
- Henneman, B., Brouwer, T. B., Erkelens, A. M., Kuijntjes, G.-J., van Emmerik, C., van der Valk, R. A., et al. (2020). Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res.* 21:gkaa1196. doi: 10.1093/nar/gkaa1196
- Henneman, B., van Emmerik, C., van Ingen, H., and Dame, R. T. (2018). Structure and function of archaeal histones. *PLoS Genet.* 14:e1007582. doi: 10.1371/journal.pgen.1007582
- Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J., and Reeve, J. N. (2014). Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genom.* 15:684. doi: 10.1186/1471-2164-15-684
- Krupovic, M., Gonnet, M., Hania, W., Ben Forterre, P., and Erauso, G. (2013). Insights into dynamics of mobile genetic elements in hyperthermophilic environments from five new *Thermococcus* plasmids. *PLoS One* 8:e49044. doi: 10.1371/journal.pone.0049044
- Krzywinski, M., Schein, J., Biro, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma* Chapter 11, Unit 11.7. doi: 10.1002/0471250953.bi1107s32
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Larsen, S. P., Bowerman, S., and Luger, K. (2021). Archaea: the final frontier of chromatin. *J. Mol. Biol.* 433, 166791. doi: 10.1016/j.jmb.2020.166791
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. doi: 10.1038/38444
- Mattiroli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart, B. W., et al. (2017). Structure of histone-based chromatin in Archaea. *Science* 357, 609–612. doi: 10.1126/science.aaj1849
- Mochizuki, T., Sako, Y., and Prangishvili, D. (2011). Provirus induction in hyperthermophilic archaea: characterization of *Aeropyrum pernix* spindle-shaped virus I and *Aeropyrum pernix* ovoid virus 1. *J. Bacteriol.* 193, 5412–5419. doi: 10.1128/JB.05101-11
- Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J. P., Reeve, J. 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
- Nishida, H., and Oshima, T. (2017). Archaeal histone distribution is associated with archaeal genome base composition. *J. Gen. Appl. Microbiol.* 63, 28–35. doi: 10.2323/jgam.2016.07.003
- Ortmann, A. C., Brumfield, S. K., Walther, J., McInnerney, K., Brouns, S. J. J., van de Werken, H. J. G., et al. (2008). Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with sulfolobus turreted icosahedral virus. *J. Virol.* 82, 4874–4883. doi: 10.1128/jvi.02583-07
- Pan, M., Santangelo, T. J., Čuboňová, L., Li, Z., Metangmo, H., Ladner, J., et al. (2013). *Thermococcus kodakarensis* has two functional PCNA homologs but only one is required for viability. *Extremophiles* 17, 453–461. doi: 10.1007/s00792-013-0526-8
- Peeters, E., Driessen, R. P. C., Werner, F., and Dame, R. T. (2015). The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.* 13, 333–341. doi: 10.1038/nrmicro3467
- Perte, M., Perte, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rogers, M. F., Thomas, J., Reddy, A. S. N., and Ben-Hur, A. (2012). SpliceGrapher: Detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. *Genome Biol.* 13:1186. doi: 10.1186/gb-2012-13-1-r4
- Rojeck, M., Hocher, A., Stevens, K. M., Merkschlager, M., and Warnecke, T. (2019). Chromatinization of *Escherichia coli* with archaeal histones. *Elife* 8:49038. doi: 10.7554/eLife.49038
- Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R., and Santangelo, T. J. (2019a). TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* 111, 784–797. doi: 10.1111/mmi.14191
- Sanders, T. J., Marshall, C. J., and Santangelo, T. J. (2019b). The role of archaeal chromatin in transcription. *J. Mol. Biol.* 431, 4103–4115. doi: 10.1016/j.jmb.2019.05.006
- Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E., et al. (2020). FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* 5, 545–553. doi: 10.1038/s41564-020-0667-3
- Sandman, K., and Reeve, J. N. (2000). Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. *Arch. Microbiol.* 173, 165–169. doi: 10.1007/s002039900122

- Sandman, K., and Reeve, J. N. (2001). Chromosome packaging by archaeal histones. *Adv. Appl. Microbiol.* 50, 73–99. doi: 10.1016/s0065-2164(01)50004-0
- Sandman, K., and Reeve, J. N. (2006). Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.* 9, 520–525. doi: 10.1016/j.mib.2006.08.003
- Santangelo, T. J., Cubonova, L., Skinner, K. M., and Reeve, J. N. (2009). Archaeal intrinsic transcription termination in vivo. *J. Bacteriol.* 191, 7102–7108. doi: 10.1128/JB.00982-09
- Sas-Chen, A., Thomas, J., Matzov, D., Taoka, M., Nance, K., Nir, R., et al. (2019). Dynamic RNA acetylation revealed by cross-evolutionary mapping at base resolution. *Nature* 583, 638–643.
- Sievert, C., Parmer, C., Hocking, T., and Chamberlain, S. (2017). *plotly: Create Interactive Web Graphics via 'plotly.js'. R package version 4.7.1.* Available online at: <https://CRAN.R-project.org/package=plotly>
- Soares, D. J., Sandman, K., and Reeve, J. N. (2000). Mutational analysis of archaeal histone-DNA interactions. *J. Mol. Biol.* 297, 39–47. doi: 10.1006/jmbi.2000.3546
- Soler, N., Marguet, E., Verbavatz, J. M., and Forterre, P. (2008). Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order *Thermococcales*. *Res. Microbiol.* 159, 390–399. doi: 10.1016/j.resmic.2008.04.015
- Stevens, K. M., Swadling, J. B., Hocher, A., Bang, C., Gribaldo, S., Schmitz, R. A., 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
- Tagashira, K., Fukuda, W., Matsubara, M., Kanai, T., Atomi, H., and Imanaka, T. (2013). Genetic studies on the virus-like regions in the genome of hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Extremophiles* 17, 153–160. doi: 10.1007/s00792-012-0504-6
- Wilkinson, S. P., Ouhammouch, M., and Geiduschek, E. P. (2010). Transcriptional activation in the context of repression mediated by archaeal histones. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6777–6781. doi: 10.1073/pnas.1002360107
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031

**Conflict of Interest:** AGE and AGa are employed and funded by New England Biolabs, Inc. New England Biolabs is a manufacturer and vendor of molecular biology reagents, including DNA replication and repair enzymes. This affiliation does not affect the authors' impartiality, objectivity of data generation or its interpretation, adherence to journal standards and policies or availability of data.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Sanders, Ullah, Gehring, Burkhart, Vickerman, Fernando, Gardner, Ben-Hur and Santangelo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.