The landscape of the histone-organized chromatin of Bdellovibrionota bacteria

Georgi K. Marinov[1,#], Benjamin Doughty[1], Anshul Kundaje[1,2], and William J. Greenleaf[1,3,4,5]

[1] *Department of Genetics, Stanford University, Stanford, California 94305, USA*
[2] *Department of Computer Science, Stanford University, Stanford, California 94305, USA*
[3] *Center for Personal Dynamic Regulomes, Stanford University, Stanford, California 94305, USA*
[4] *Department of Applied Physics, Stanford University, Stanford, California 94305, USA*
[5] *Chan Zuckerberg Biohub, San Francisco, California, USA*
[#] *Corresponding author*

Abstract

Histone proteins have traditionally been thought to be restricted to eukaryotes and most archaea, with eukaryotic nucleosomal histones deriving from their archaeal ancestors. In contrast, bacteria lack histones as a rule. However, in recent years histone proteins have been identified in a few bacterial clades, in particular the phylum Bdellovibrionota, and these histones have been proposed to exhibit a range of divergent features compared to histones in archaea and eukaryotes. However, no experimental functional genomic studies of the properties of Bdellovibrionota chromatin have been carried out. In this work, we map the landscape of chromatin accessibility, active transcription and three-dimensional genome organization in a member of Bdellovibrionota (a *Bacteriovorax* strain). We find that *Bacteriovorax* chromatin is characterized by preferential accessibility around promoter regions, similar to what is observed in eukaryotes with compact genomes such as yeast, and also to some archaea. As in eukaryotes, chromatin accessibility positively correlates with gene expression. Mapping active transcription through single-strand DNA (ssDNA) profiling revealed that *Bacteriovorax* promoters exhibit very strong polymerase pausing, unlike in yeast, but similar to the state of mammalian and fly promoters. Finally, the *Bacteriovorax* genome exists in a three-dimensional (3D) conformation analogous to that of other bacteria without histones, organized by the parABS system and along the axis defined by replication origin and termination regions. These results provide a foundation for understanding the chromatin biology of the unique Bdellovibrionota bacteria and the deep evolution of chromatin organization across the tree of life.

Introduction

Histones and nucleosomal chromatin are one of the defining features of the broadest divisions of life on our planet. Nearly all eukaryotes[6] package their chromatin around nucleosomal particles, composed of two dimerized tetramers of the four core histones H2A, H2B, H3 and H4, and histones are the most conserved proteins in their genomes[7], especially when it comes to the key residues playing a vital role in the so called "histone code"[8] (and through it in all aspects of chromatin biology – transcription, gene regulation, DNA repair, DNA replication, mitosis and meiosis, and many others), which are almost universally invariant across all eukaryotes.

Eukaryotic histones trace their ancestry to Archaea[9], which is the half of the prokaryote diversity from within which eukaryotes appear to have emerged, and where the informational processing systems of eukaryotes are thought to derive from[10–15]. However, archaeal histones are quite different from those of eukaryotes. While they all contain the same cannonical histone fold domain[16], consisting of three alpha helices[17] dimerizing with another histone molecule, archaeal histones only contain the histone fold domain[18,19] while those of eukaryotes have long tails (that are the primary sites of post-translational modifications). Another major difference is that unlike the eukaryotic octamer nucleosome, archaeal histone dimers can oligomerize into so called "hypernucleosomes"[20–23], which are stacks of individual histone dimers, and can be of variable length.

Bacteria are the other main branch of prokaryotes and are generally thought to not have histones. However, recent phylogenomic and experimental studies have revealed that some bacteria in fact do possess histone proteins[24,25].

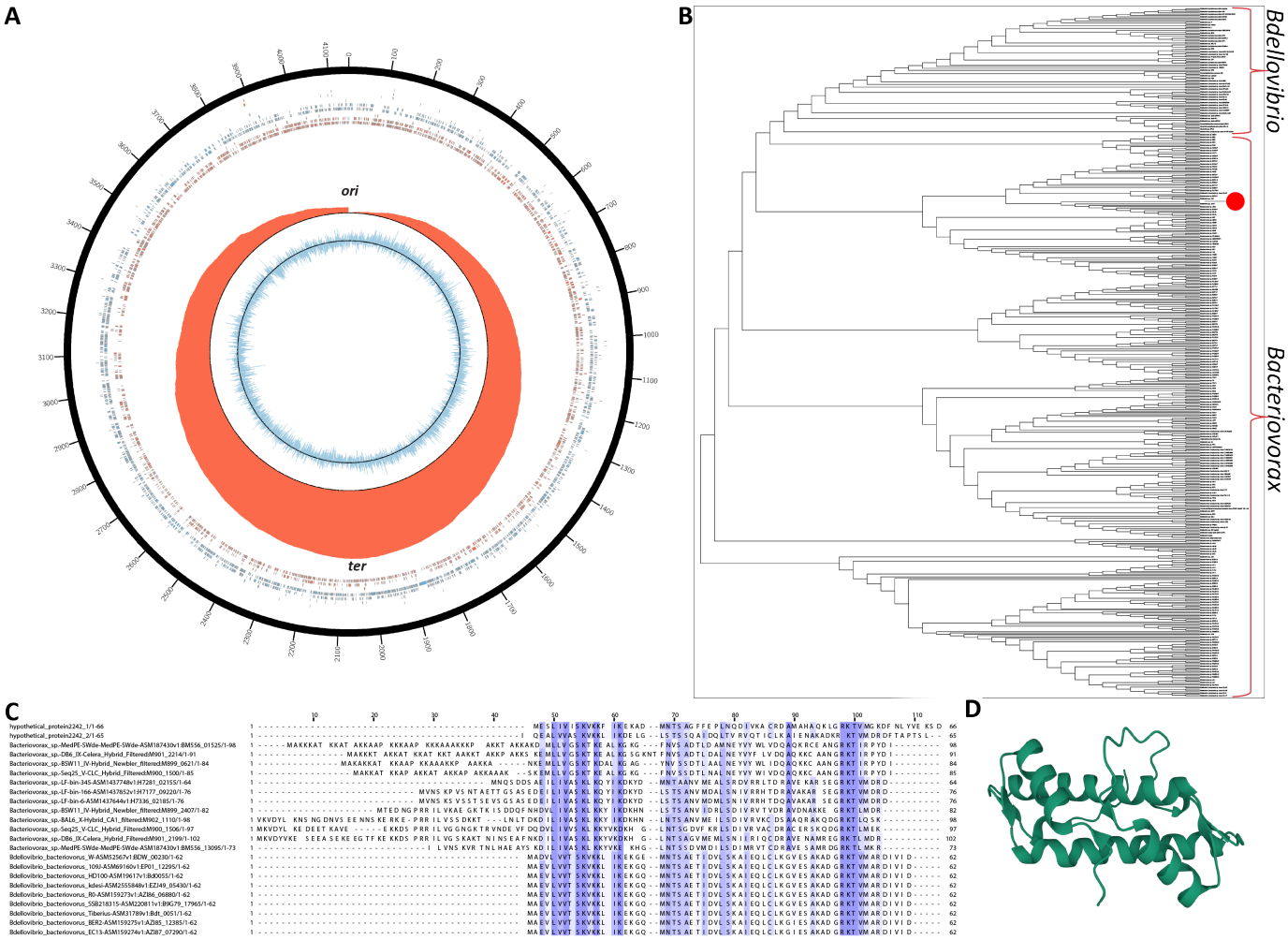


Figure 1: Genome assembly and annotation of *Bacteriovorax* sp. ICPB 3264 [H-I A3.12]. (A) Circos[1] plot of the *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] chromosome. Forward- and reverse-strand protein coding genes are shown in red and blue respectively in the inner tile layer; ribosomal RNAs and tRNAs are shown in the outer tile layers. The red histogram shows the cumulative GC-skew[2]; with the inflection points corresponding to the replication origins and and termini[3]. (B) Maximum likelihood 16S rRNA tree (generated using RAxML-NG[4]) of members of the Bdellovibrionota phylum and the phylogenetic location of strain ICPB 3264 [H-I A3.12]. (C) Multiple sequence alignment of histone proteins identified in select *Bacteriovorax* and *Bdellovibrio* genomes. In contrast to the singlet histone folds in *Bdellovibrio*, the histone gene in *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] is a histone doublet; the two histone folds are shown separately (top two rows). (D) Predicted[5] protein structure of the histone gene in *Bacteriovorax* sp. ICPB 3264 [H-I A3.12].

Most notably in the bacterium *Bdellovibrio bacteriovorus* and other members of the Bdellovibrionota phylum, histones are both major in abundance and essential for viability components of the nucleoid[26,27]. Bdellovibrionota histones have been shown to bend DNA as dimers and in a sequence-independent manner[27], while other studies have suggested that they exhibit yet another major deviation from the conventional eukaryote state of nucleosomal architecture by binding to DNA head-on, coating it, as opposed to the situation in eukaryotes and archaea, where DNA is wrapped around histones[26].

These observations pose a long list of tantalizing questions about the physical organization of the genomes of these histone-possessing bacteria, about the relationship between chromatin structure and gene expression, and about how such features compare to those of archaea. However, while archaea have received some experimental attention in the past decade[28–31], these properties have not been assayed so far with functional genomic tools in the bacterial clades that have histones.

In this study, in order to begin addressing these outstanding issues, we have mapped chromatin accessibility using ATAC-seq[32] (Assay for Transposase-Accessible Chromatin using sequencing),transcriptional activity us-

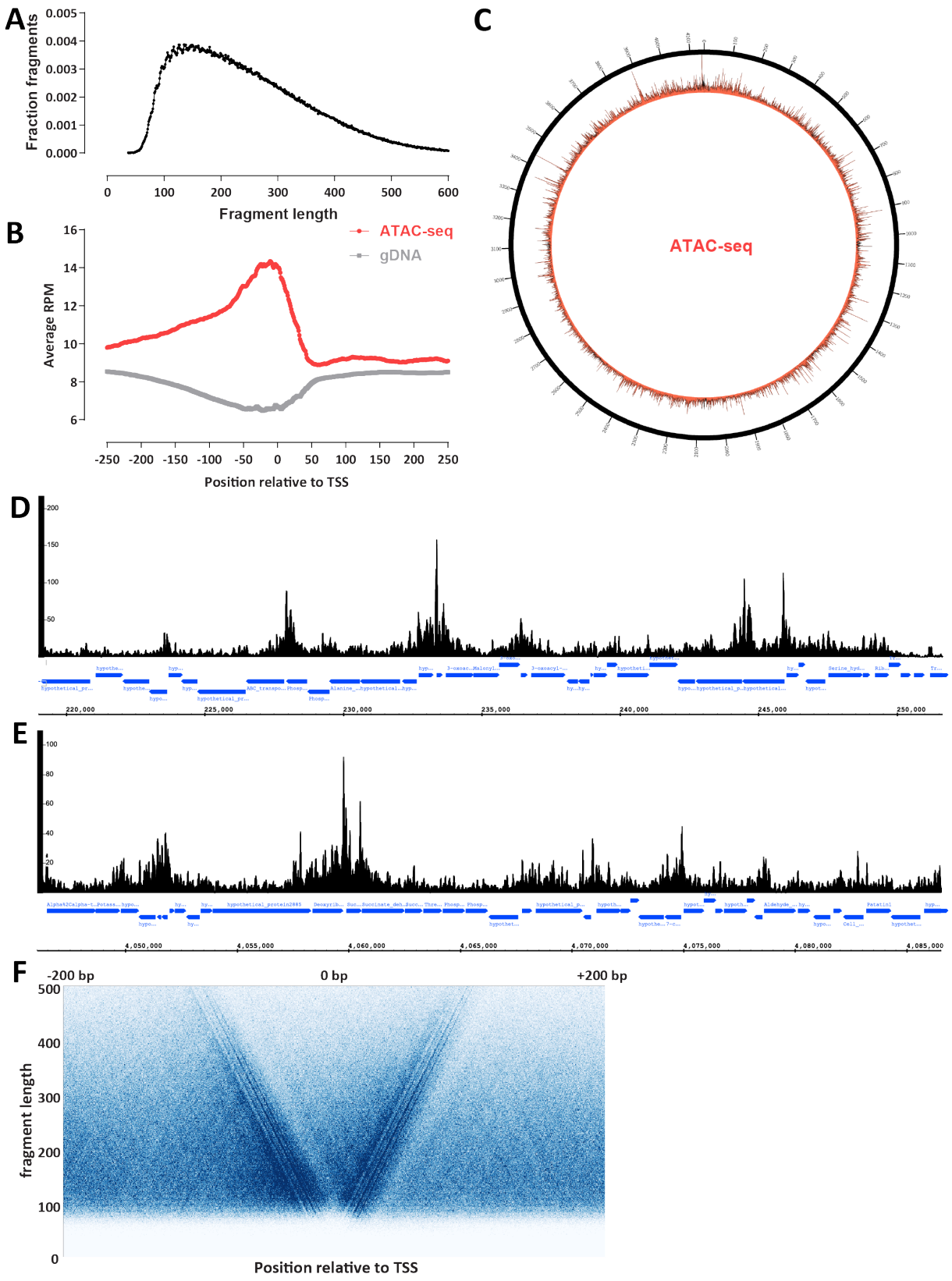


Figure 2: The accessible chromatin landscape in *Bacteriovorax* . (A) ATAC-seq fragment length distribution. (B) ATAC-seq metaprofile around annotated *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] TSSs. (C) Circos plot of the global accessibility distribution along the *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] chromosome. (D-E) Representative snapshots of local ATAC-seq signal distribution in *Bacteriovorax* (exponentially growing culture). (F) V-plot of ATAC-seq fragment distribution around annotated *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] TSSs.

ing KAS-seq[35] (Kethoxal-assisted single-stranded DNA sequencing), and three-dimensional genome organization using Hi-C[36] in a member of the Bdellovibrionota phylum (a *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] strain). Our data reveals a chromatin accessibility landscape reminiscent of that of eukaryotes with highly compact genomes, with preferentially accessible promoter regions, strong polymerase pausing at promoters, positive correlation between promoter accessibility and active transcription rates, and conventional for bacteria three-dimensional organization. We discuss these properties in the broad context of the diversity of life on our planet.

Results

Genome assembly and annotation of *Bacteriovorax* sp. ICPB 3264 [H-I A3.12]

We first set out to map the chromatin accessibility landscape in *Bdellovibrio bacteriovorus* using ATAC-seq, given that *Bdellovibrio* histones were the ones characterized by recent studies of bacterial histones[26] and *Bdellovibrio bacteriovorus* is the most well-studied Bdellovibrionota representative. However, a major technical challenge to the application of ATAC-seq to *Bdellovibrio* is presented by the unique biology of these bacteria. *Bdellovibrio* is the most famous example of a predatory prokaryote[37–39] – it feeds by attaching onto the cell wall of gram-negative bacteria, creating an opening into it, then inserting itself into the periplasmic space between the inner and outer bacterial membranes, breaking down the host cell, subsequently undergoing polyploid division, and eventually lysing the remnants of the host. This means that *Bdellovibrio bacteriovorus* is usually grown together with another prey species, which are bacteria without histones.

This is problematic in terms of applying ATAC-seq to *Bdellovibrio* because the basic principle behind ATAC-seq is the strong preference of the Tn5 transposase for open chromatin not protected by nucleosomes[32]. In mammalian cells, this can result in extreme overrepresentatin of mitochondrial DNA sequences in final libraries if mitochondria have not been properly filtered out during the nuclei isolation procedure. In a mixture of bacteria without and with histones, if the histones of the latter are strongly protecting DNA from Tn5 insertion, ATAC-seq libraries could easily consist almost entirely of fragments from the prey rather than *Bdellovibrio*.

Fortunately, prey-independent strains of *Bdellovibrio* are also available, allowing axenic growth in media free of other bacteria. We obtained "*Bdellovibrio bacteriovorus*" strain ICPB 3264 [H-I A3.12] from ATCC and carried out ATAC-seq on it. However, almost no reads from these libraries aligned to any of the previously sequenced Bdellovibrionota strains available in the NCBI databases (neither *Bdellovibrio* nor strains from *Bacteriovorax*, which is the other major division of the phylum).

We thus carried out *de novo* genome sequencing of the strain we worked with, using a combination of nanopore long reads and short Illumina reads (see the Methods section for details). This resulted in a single contig of length 4,148,738 bp (∼667× coverage) predicted to encode (see Methods for details) 4,127 protein coding genes (Figure 1A).

Analysis of potential DNA modifications revealed the existence of 5-methylcytosine methylation in a CpGGCC context. This is likely associated with a restriction modification (R-M) system[40] in *Bacteriovorax*.

We then established the precise phylogenetic positioning of the strain by analyzing available 16S rRNA sequences for Bdellovibrionota strains. The phylum Bdellovibrionota[41] includes three major clades/classes – Bacteriovoracia, Bdellovibrionia, and Oligoflexia. Bacteriovoracia includes the *Bacteriovorax*, *Halobacteriovorax* and *Peredibacter* genera while Bdellovibrionia features *Bdellovibrio* and *Pseudobdellovibrio*. Our phylogenetic analysis points to strain ICPB 3264 [H-I A3.12] beloning to the *Bacteriovorax* genus, being most closely related to *Bacteriovorax* sp. B1T4-F (Figure 1B), even though it was originally labeled as a "*Bdellovibrio*" strain. We refer to it as *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] or simply *Bacteriovorax* in the subsequent text.

Finally, we examined putative *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] histone genes in order to confirm their presence in its genome, and found one histone protein to be encoded in our assembly. Unlike *Bdellovibrio* histone genes, which encode histone singlets that later form dimers, in *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] the histone gene encodes a histone doublet that presumably functions as a dimer on its own (Figure 1C–D).

ATAC-seq reveals the accessible chromatin landscape in *Bacteriovorax*

In order to map the chromatin accessibility landscape in *Bacteriovorax*, we adapted the previously described bacterial ATAC (bac-ATAC) protocol[42], which involves crosslinking with 1% formaldehyde and permeabilization of the cell wall with lysozyme (see the Methods section for details).

In metazoans, ATAC-seq libraries display a clear nucleosomal signature, with a strong subnucleosomal peak ($\leq$120 bp) followed by a robust mononucleosomal ($\sim$ 150 bp) peak, a weaker dinucleosomal peak, and so on. In *Bacteriovorax*, we do not observe a nucleosomal protection pattern but only a single broad fragment peak centered between 100 bp and 200 bp (Figure 2B). Thus *Bacteriovorax* histones do not appear to confer the same arrayed local protection to DNA as eukaryotic ones.

Accessibility is centered around the promoter regions between genes (Figure 2B-E). We observe transcription start sites (TSS) scores (indicating the level of enrichment over promoter regions relative to flanking regions; see the Methods section and previous detailed discussions[43]) in the 1.35-

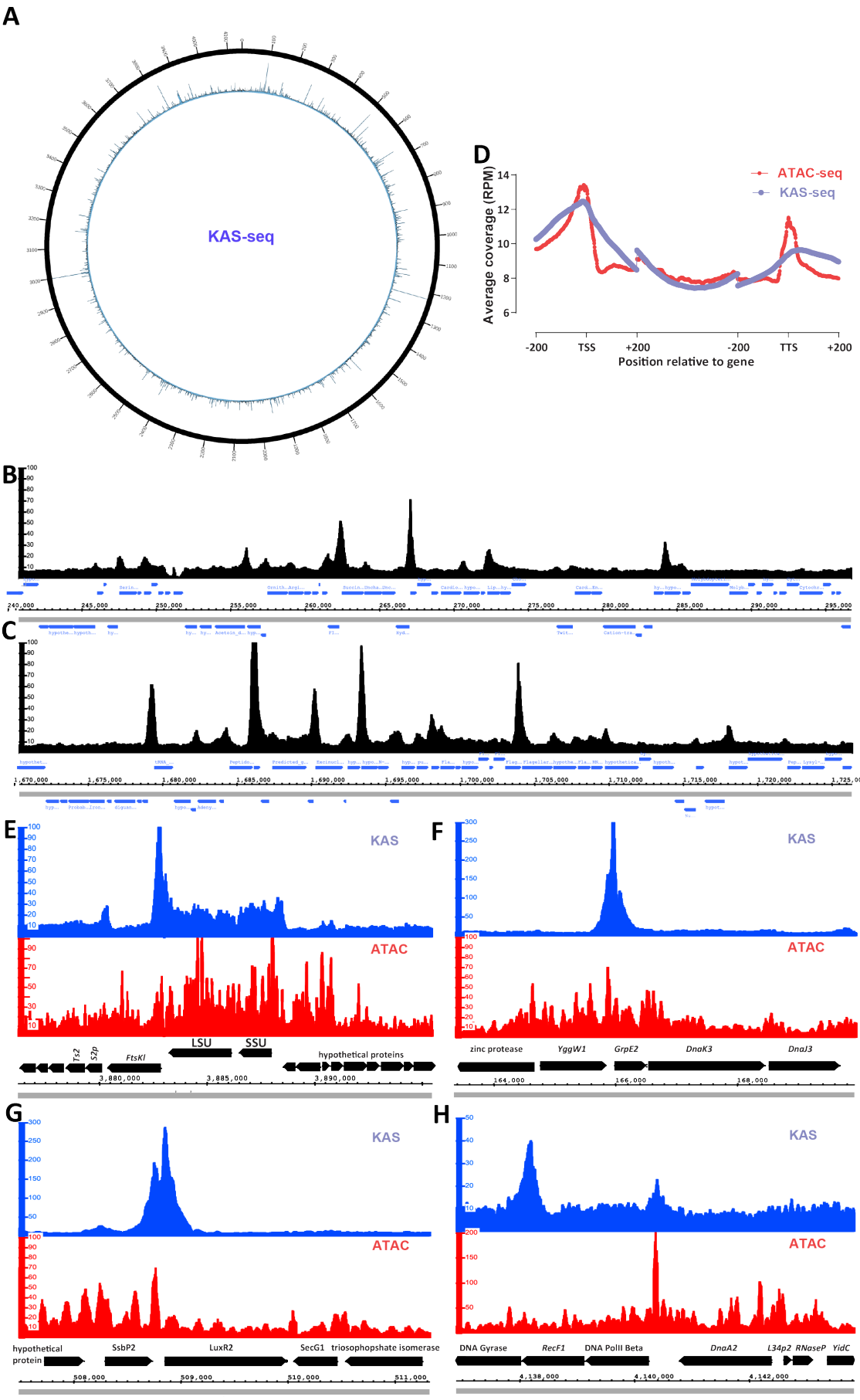


Figure 3: The ssDNA and active transcription landscape in *Bacteriovorax* . (A) Circos plot of the global KAS-seq signal distribution along the *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] chromosome. (B-C) Local KAS-seq signal distribution in *Bacteriovorax* shows strongly localized peaks associated with promoter regions (D) Metaplot of KAS-seq and ATAC-seq signal distribution along *Bacteriovorax* genes. (E) KAS-seq and ATAC-seq profiles around the ribosomal RNA locus of *Bacteriovorax*. (F) KAS-seq and ATAC-seq profiles around the heat shock protein *GrpE2* gene. (G) KAS-seq and ATAC-seq profiles around the *LuxR2* transcriptional activator gene. (H) KAS-seq and ATAC-seq profiles near the origin of replication of the *Bacteriovorax* chromosome.

1.40 neighborhood, which is comparable to what we had previously found in the archaeon *Haloferax volcanii*[31] and also to the lower end what is usually seen in eukaryotes with highly compact genomes such as yeast[44]. In contrast, naked genomic DNA controls exhibit slight depletion around TSS (Figure 2B), confirming the non-artefactual origin of the observed enrichment. Manual inspection of ATAC-seq profiles along the genome confirms these global observations, but also shows frequent cases of elevated accessibility extending into gene bodies (Figure 2D-E).

We also note that a previous ATAC-seq study of conventional bacteria that do not possess histones (*Caulobacter crescentus*) reported the existence of broad domains of elevated and decreased accessibility spanning hundreds of kilobases[42]. This is not readily observed in *Bacteriovorax* (Figure 2C), just as it is not observed in the archaeon *Haloferax volcanii*.

Finally, we asked whether patterns of nucleosomal positioning are observed around promoters using V-plot analysis[45]. With the caveat that we do not yet have absolutely precise TSS annotations for *Bacteriovorax*, we do not observe strongly positioned nucleosomes around its TSSs (Figure 2F).

The ssDNA and active transcription landscape in *Bacteriovorax*

Next, we mapped the landscape of active transcription in *Bacteriovorax*. To this end we employed the KAS-seq assay[35], which strongly and specifically enriches for ssDNA structures by labeling exposed guanine (G) bases with N_3-kethoxal, followed by biotin labeling through a click reaction, fragmentation of DNA, and streptavidin pulldown (see the Methods section for details). While KAS-seq labels all ssDNA structures (G-quadruplexes, replication intermediates, etc.), the most abundant by a large distance source of ssDNA in the cell are transcriptional bubbles associated with RNA polymerases directly engaged with DNA, both paused and actively elongating[35], thus KAS-seq datasets are an excellent way to map active transcription in a simple and efficient way.

As with ATAC-seq, we do not observe broad large-scale domains of active transcription along the *Bacteriovorax* chromosome in our KAS-seq datasets (Figure 3A). The global landscape is largely uniform, punctuated by strong localized peaks.

Strikingly, examination of local browser tracks (Figure 3B-C) revealed these peaks to be associated with promoters, indicating the existence of very strong promoter pausing in *Bacteriovorax*. Polymerase pausing and its later controlled release are decisive steps in gene regulation in many eukaryotes[46], classically having been first described in the fruit fly *Drosophila melanogaster*[47–50], and also being widespread in mammals[51,52], most, though not all metazoans, and some plants, but curiously absent in yeast and *Arabidopsis thaliana*[53]. Bacteria do exhibit sequence-dependent polymerase pausing but this has been primarily reported in the context of transcription elongation and termination[54–59] and *E. coli* promoters are not characterized by the typical peaks associated with pausing in global run-on GRO-seq datasets[53]. The strong KAS-seq peaks we observe are not due to transcription termination-associated pausing because they are found in all possible orientations of gene pairs, including between two divergent promoters (Figure 3B-C).

Of note, we had previously observed promoter pausing using KAS-seq in the euryarchaeon *Haloferax volcanii*[31]. *Bacteriovorax* promoters are even more accentuated in KAS-seq profiles than in *Haloferax* – promoter KAS signal peaks at ∼12 RPM (Reads Per Million mapped reads) over *Bacteriovorax* promoters compared to a through at ∼10 RPM in intergenic space and ∼8 over gene bodies (Figure 3B-D); in *Haloferax* these values are ∼11, ∼10 and ∼8 RPM, respectively[31].

The most strongly transcribed (as assessed by KAS-seq levels over the gene body) genes in *Bacteriovorax* are the two ribosomal RNAs (Figure 3E). Notably, these do not exhibit a KAS-seq peak in their promoter (although a very strong such peak is located downstream of the ribosomal RNA operon). This is also one of the most accessible over gene bodies loci in the genome as measured by ATAC-seq; this points to an analogous situation to the one well known from yeast and other eukaryotes. Yeast rDNA exists in two states – a transcriptionally inactive chromatinized condition on one hand, and a highly transcriptionally active, almost entirely devoid of histones state[44,60,61] on the other. These observations also contrast with those from ATAC-seq for *Caulobacter crescentus*, which does not posses histones, and in whose genome ribosomal RNA clusters were reported to be one of a handful of highly transposase-inaccessible transcribed regions[42] (HINTs).

Other genes where particularly strong putative paused RNA polymerase peaks are observed include the *GrpE2* heat shock protein (Figure 3F), a *LuxR2* transcriptional activator gene (Figure 3G), the chaperone protein ClpB ATP-dependent unfoldase (Supplementary Figure 1A), the RNA Polymerase Sigma Factor RpoH1 (Supplementary Figure 1B), and others. For each of these proteins, especially the heat shock ones, it is plausible that polymerase pausing is a regulatory mechanism specifically designed to enable their very rapid activation, in a manner analogous to heat shock genes in *Drosophila* where the phenomenon was originally described[47,48].

The strongest ATAC-seq peak in the genome is found in an unusually large integenic space without any annotated genes located between replication-related genes and near the origin of replication (Figure 3H). This locus exhibits only a modest ssDNA peak, and may possibly be associated with replication initiation.

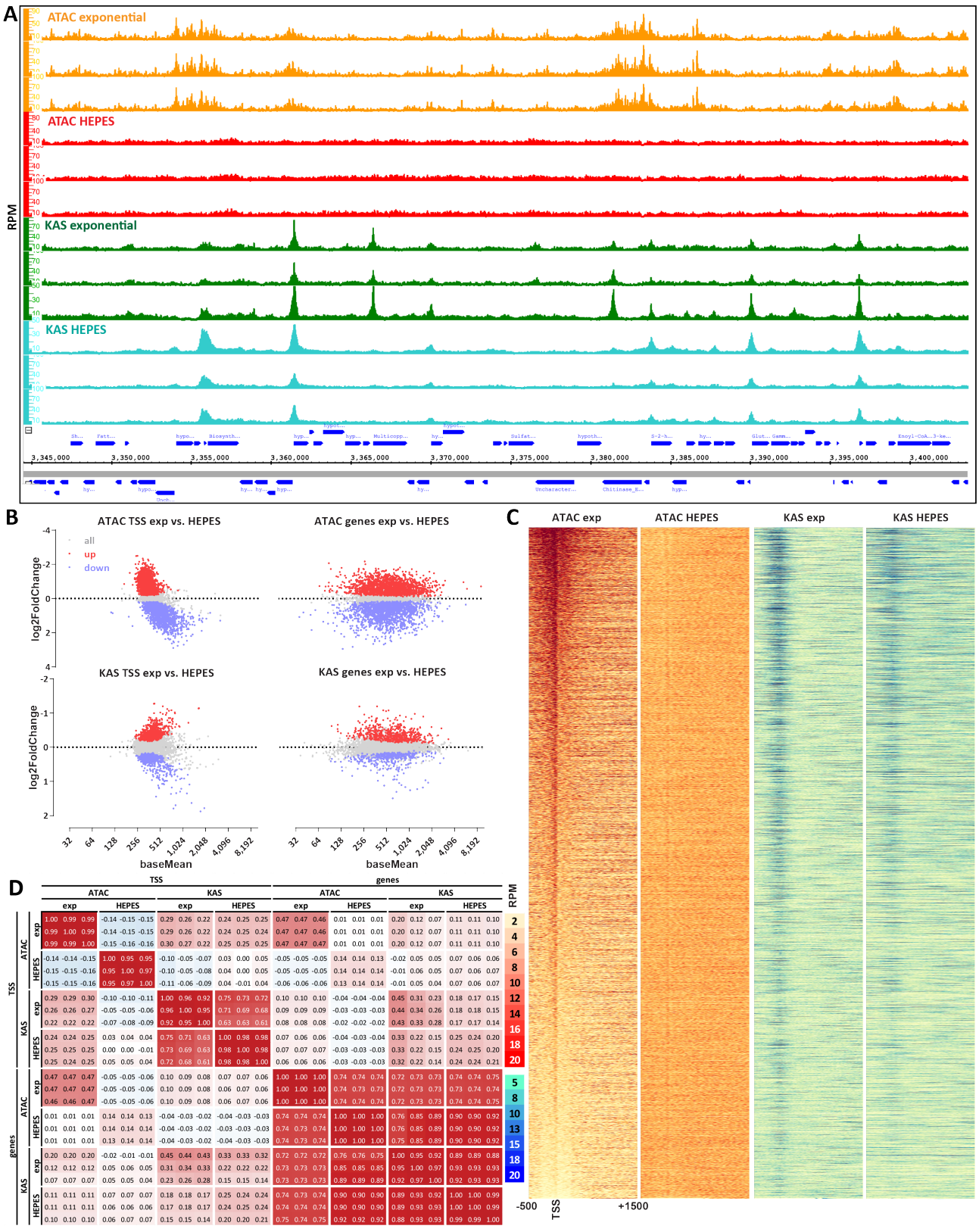


Figure 4: Relationship between chromatin accessibility and transcriptional activity in *Bacteriovorax*. (A) Starvation of *Bacteriovorax* cells induces changes in the chromatin accessibility and active transcription landscapes. Show are three independent KAS-seq and ATAC-seq replicates for each condition. (B) Differential accessibility and KAS levels between exponentially growing and starved (HEPES) cells. (C) Global ATAC-seq and KAS-seq profiles over *Bacteriovorax* genes under exponentially growing and starvation conditions. (D) Correlation between ATAC-seq and KAS-seq signal over TSSs and gene bodies.

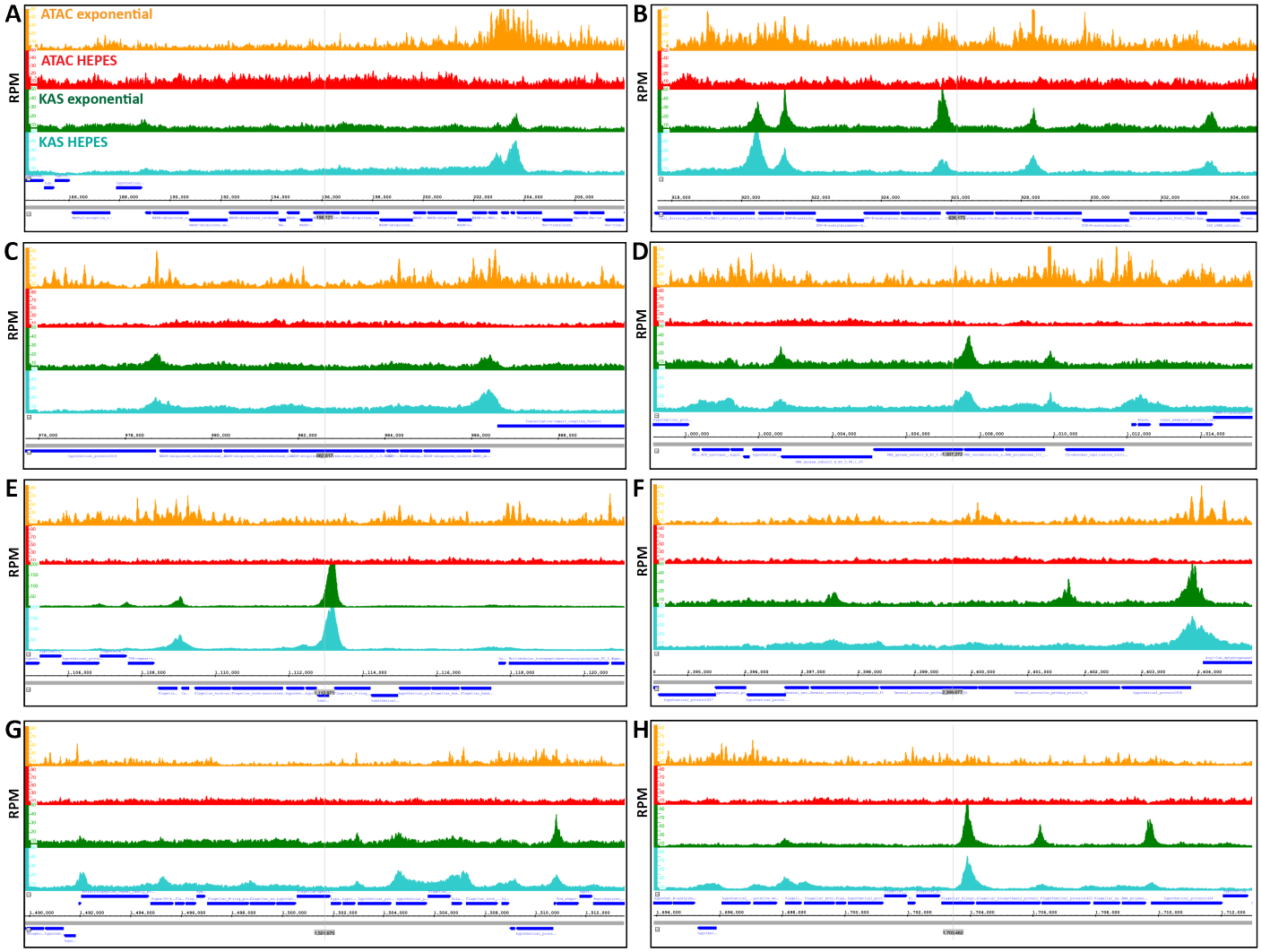


Figure 5: Chromatin accessibility and active transcription levels in *Bacteriovorax* operons . (A-H) Genome browser snapshots of ATAC-seq and KAS-seq levels in exponentially growing and starved (HEPES) cells for eight different *Bacteriovorax* operons.

Chromatin accessibility correlates with transcriptional activity in *Bacteriovorax*

Our next step was to characterize the relationship between chromatin accessibility and transcriptional activity in *Bacteriovorax* as well as their dynamics upon large-scale gene expression perturbations. Genome-wide gene expression dynamics in these organisms has not been studied extensively in the literature, but previously it was reported that starvation of *Bdellovibrio* cells for 4 hours in HEPES buffer does result in major changes in gene expression[62].

We carried out ATAC-seq and KAS-seq in multiple replicates from the same cells and at the same time in both exponentially growing and HEPES-starved cultures, and indeed observed large-scale changes in transcriptional activity and chromatin accessibility (Figure 4A-C). Curiously, chromatin accessibility around promoters largely disappeared in starved cells, even though transcriptional activity was either unaffected at many promoters or shifted markedly in others but overall preserved similar global properties to exponentially growing cells (except for a slight shift from promoters towards presumed elongation over gene bodies for a number of genes; Figure 4C). We identified 555/378 gene bodies and 488/607 TSSs with respectively statistically decreased/increased KAS-seq signal.

Previously we found no correlation between chromatin accessibility and transcriptional activity in the archaeon *Haloferax volcanii*[31]. In contrast, ATAC-seq signal both over promoter regions and gene bodies correlates positively with KAS signal, again over both promoter regions and gene bodies (Figure 4D).

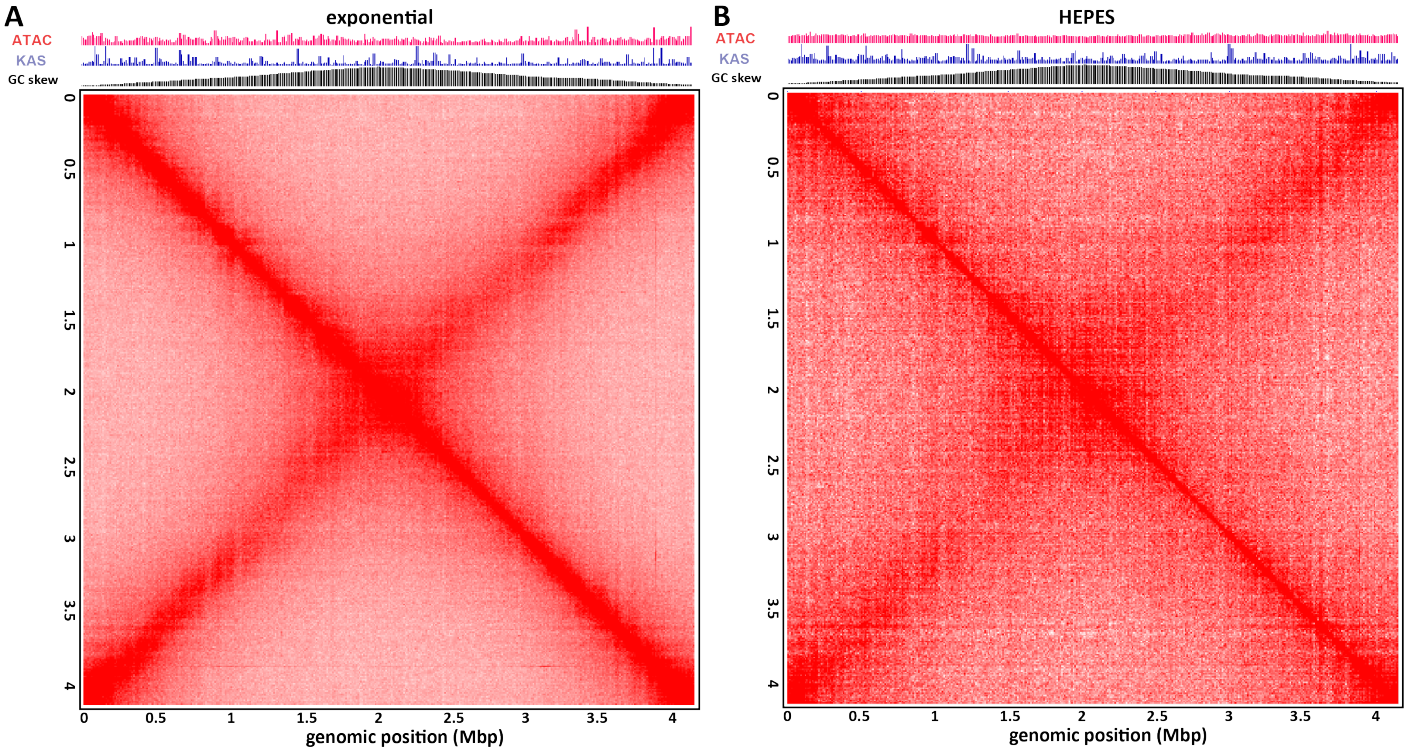


Figure 6: Three-dimension organization of the the *Bacteriovorax* sp. ICPB 3264 [H-I A3.12] chromosome. (A) Hi-C map of exponentially growing *Bacteriovorax* cells. (B) Hi-C map of starved (HEPES) *Bacteriovorax* cells.

Frequent independent transcription initiation/regulation of individual genes within *Bacteriovorax* operons

In both bacteria and archaea functionally related genes are often organized into operons[63] that are transcribed together. However, evidence has accumulated over the years that certain genes within operons may be transcribed and regulated separately from the whole operon[64–67]. KAS-seq and ATAC-seq data for the archaeon *Haloferax* also demonstrated this phenomenon on a genome-wide scale – individual promoters, discernible by strong ATAC-seq and/or KAS-seq peaks inside operons, are frequently observed in that organism, as are obviously different KAS-seq levels over different genes within a single operon[31].

Bdellovibrionota are bacteria, and we aimed to determine whether similar unexpected complexity in the functional organization of operons exists in their genomes too. To this end we identified cases of clear operons (i.e. colinear genes obviously belonging to the same functional group) and examined KAS-seq and ATAC-seq profiles in both conditions that we assayed (Figure 5). We observed multiple

	clade	histones	histone-based chromatin	genome size	gene length		intergenic space		ATAC TSS score	promoter pausing
					mean	median	mean	median		
Bacteriovorax sp.	Bacteria	prokaryotic	✓	4,148,738	939	785	94	65	1.35-1.4	✓
Caulobacter crescentus		×	×	4,042,929	944	800	128	97	low	?
Haloferax volcanii	Archaea	prokaryotic	?	4,012,900	833	740	167	108	1.35-1.4	✓
Sulfolobus islandicus		×	×	2,522,992	823	722	194	77	1.1	?
Saccharomyces cerevisiae	Eukaryotes	H2A, H2B, H3, H4	✓	12,157,105	1,300	1,019	498	347	1.4-2	×
Drosophila melanogaster		H2A, H2B, H3, H4	✓	143,726,002	5,753	1,709	3,597	644	1.6-2.3	✓
Homo sapiens		H2A, H2B, H3, H4	✓	3,099,750,718	67,046	26,780	113,388	20,430	10-25	✓

Figure 7: Comparison of the chromatin landscape across the deep organismal diversity. Shown is the presence of histones, histone-based chromatin, the basic genomic properties, i.e. genome size, gene length, gene density/intergenic space size, and the extent of concentration of ATAC-seq signal around promoters ("TSS score", see the Methods section for details).

cases of both internal KAS-seq and ATAC-seq peaks inside operons (e.g. Figure 5B and Figure 5E). While internal KAS-seq peaks could be associated with polymerase pausing due to regulation of transcriptional elongation, RNA processing or cotranscriptional translation events, internal ATAC-seq peaks most likely correspond to independent promoters.

Three-dimensional organization of the *Bacteriovorax* genome

Finally, we characterized the 3D organization of the *Bacteriovorax* genome. We employed a modification of the Hi-C[36] assay, using multiple restriction 4-cutter enzymes to digest crosslinked *Bacteriovorax* chromatin in order to improve resolution in its very small and compact genome (see the Methods section for details).

Hi-C has been previously applied to study the physical organization of several bacterial and archaeal strains, such as originally *Caulobacter crescentus*[68,69] and later multiple others[70–77]. These studies have revealed two main typical features of bacterial chromosomes. In *Caulobacter* and most other bacteria, Hi-C maps exhibit a prominent pattern of increased interactions perpendicular to the main diagonal and connecting the origin and the terminus of replication, reflecting a chromosome configuration driven by SMC/condensin complexes loaded at centromere-like parS sites near replication origins[78]. The parABS system[79–82] plays a major role in this process, with most bacteria having one or multiple highly conserved[83] parS sites near their origin of replication. The second common major feature are self-interacting chromosomal interaction domains (CIDs), areas of increased local contact frequency corresponding to plectonemes arising as a result of transcription-induced supercoiling[68].

Hi-C maps of archaea differ considerably – *Sulfolobus*, which does not have histones, appears to possess three distinct origins of replication resulting in a much more complicated chromosomal configuration, and does not display CIDs[84], while *Haloferax*, which has histones (although it is not clear to what extent they package the genome), was reported to exhibit CIDs but not the perpendicular to the main diagonal feature typical to bacteria[85].

In contrast, *Bacteriovorax*, although it does possess histones, exhibits a typical large-scale bacterial organization of its chromosome, with very strong cross-diagonal interactions connecting the replication origin and termination regions (Figure 6A). However, we do not observe clear evidence for plectonemic CID-like structures in our maps. HEPES-starved cells exhibit decreased strength of the cross-diagonal feature (Figure 6B), likely reflecting much lower replication activity.

Discussion

That bacteria with histone-based chromatin exist represents one of the most surprising and exciting discoveries in chromatin biology in recent years, coming after the assumption that histones are restricted to archaea and eukaryotes had reigned for many decades. Our study, carried out in a novel *Bacteriovorax* sp. strain whose genome we assembled and annotated *de novo*, presents the first direct measurements of chromatin accessibility, the distribution of active transcription and ssDNA, and three-dimensional genome conformation in such bacteria, and it allows us to chart some of the broadest trends in chromatin organization across the tree of life (Figure 7).

Like the euryarchaeon *Haloferax*, and similar to eukaryotes with nucleosomal chromatin, *Bacteriovorax*'s chromatin is characterized by preferentially accessible promoter regions. The relative level of promoter accessibility is comparable to that of *Haloferax*, slightly lower than yeast, and considerably lower than metazoans with much larger genomes. Likely this is a reflection of high transcriptional activity in extremely compacted genomes, i.e. actively transcribed regions temporarily lose the protection from transposase insertion conferred by histones or other packaging proteins, and the density of actively engaged polymerase molecules is higher in such cases than it is in mammals with highly bloated genomes. We see some more direct evidence for this phenomenon in the case of the rDNA operon in *Bacteriovorax*, which is the most broadly accessible gene in the genome, similar to the situation with rDNA in yeast.

Common with eukaryotes, but unlike the euryarchaeon *Haloferax*, *Bacteriovorax* exhibits positive correlation between chromatin accessibility and active transcription, suggesting that displacing histones from promoters might play a regulatory role in this organism. There are two observations that complicate such an interpretation. First, we do not observe a protection signature in the ATAC-seq fragment length distribution analogous to what is seen in eukaryotes and also in some archaea with histone-based chromatin[86]; it is thus not entirely clear how exactly *Bacteriovorax* histones physically protect DNA and bind to it (somewhat different models have been proposed in the past year[26,27]). Second, in starved cells we observe loss of ATAC-seq enrichment over promoters but not of actively transcribing polymerases; this can be interpreted in two ways – either promoters close upon strong stress signals, but *Bacteriovorax* histones are not strongly inhibitory to polymerase activity, or, alternatively, histones are lost from chromatin, but this does not strongly affect polymerase engagement with DNA, even though histones appear to be essential in these organisms. Single-molecule mapping of chromatin accessibility[33,34,44] and examination of a much broader array of conditions can be expect to shed light on these questions.

Surprisingly, we also observe evidence for very strong polymerase pausing over *Bacteriovorax* genes, in contrast to what has been reported for bacteria such as *E. coli*[53], and similar to what is seen in many (though not all) eukaryotes. Thus regulation of productive transcriptional elongation might be a key mechanism of modulating gene expression in bacteria with histone-based chromatin.

We also find evidence for independent regulation of and transcription initiation from internal promoters inside *Bacteriovorax* operons. This is similar to what is observed in the euryarchaeon *Haloferax*.

Finally, the *Bacteriovorax* genome exhibits the typical for bacteria three-dimensional organization centered on the axis defined by the replication origin and terminus. However, it does not display clear plectonemic CIDs. This is an interesting observation that might further illuminate the mechanisms of formation or lack of such domains. Strong plectonemic CIDs are rare in eukaryotes with one notable exception – dinoflagellates, which display the largest and most pronounced such domains of any organism assayed so far, likely thanks to their unique genomic organization featuring unidirectional gene arrays many hundreds of kilobases long (thus generating extreme levels of supercoiling stress) and the also unique to them loss of histones as a main packaging component; it has been thus hypothesized that the frequent and strong interactions between nucleosomes in other eukaryotes overrides the formation of plectonemes driven by transcription-induced supercoiling, but in dinoflagellates that effect is unmasked[87]. Conversely, if analogous such interactions are introduced in groups that normally do not feature histones and exhibit supercoiling domains, such as bacteria, it may be expected for CID-like domains to disappear. We do not observe clear CIDs in *Bacteriovorax*, but they are also not always robust in all bacteria, and not seen in the non-histone carrying crenarcheaote *Sulfolobus* while having been reported from *Haloferax*. However, more recent evidence points to *Haloferax* possessing histone genes but not actually having histone-based chromatin[88,88–90], in which case some other chromatin protein must be responsible for the high protection from Tn5 insertion measured over its genome, and other chromatin proteins might also be actively shaping the physical genome of other archaeons and bacteria. Further expanding the coverage of physical genome mapping over the deep diversity in the tree of life can be expected to resolve these and many other questions.

Methods

Except where explicitly indicated otherwise, data was processed using custom-written Python scripts (`https://github.com/georgimarinov/GeorgiScripts`)

Cell culture

We obtained the strain designated as *Bdellovibrio bacteriovorus* "ICPB 3264 [H-I A3.12]" from ATCC (`https://www.atcc.org/products/25637`). As discussed in the main text, subsequent analysis revealed that this strain does not in fact belong to the *Bdellovibrio bacteriovorus* clade, but is instead a *Bacteriovorax* strain that for which sequence data did not previously exist.

The cells were grown in ATCC medium 526 (10.0 g Peptone, 3.0 g Yeast extract, 1.0 L distilled water, autoclaved at 121 °C) at 30 °C, except where otherwise indicated for the HEPES starvation condition (in which case cells were centrifuged down and the media was replaced with HEPES buffer supplemented with 2 mM $CaCl_2$ and 4 mM $MgCl_2$ as previously described[62]).

Genome sequencing

Genomic DNA (gDNA) was isolated using the NEB Monarch Genomic DNA Purification Kit (Cat # T3010).

For Illumina sequencing, genomic DNA libraries were generated using Tn5 transposition as previously described[31]. Briefly, ∼200 ng DNA gDNA were used, with the volume increased to 22.5 μL using ultrapure H_2O. Transposition was carried out with 22.5 μL gDNA, 25 μL 2×TD buffer (20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20% Dimethyl Formamide), and 2.5 μL Tn5, incubating at 37 °C for 15 minutes. Transposed DNA was immediately purified using the Qiagen MinElute PCR Purification Kit (Cat # 28004), with the reaction stopped with 250 μL PB buffer, and elution in 10 μL EB buffer. Final library amplification was carried out by mixing the 10 μL eluate, 10 μL H_2O, 2.5 μL i5 primer, 2.5 μL i7 primer, and 25 μL NEBNext High-Fidelity 2× PCR Master Mix, using the following thermocycler program: 3 minutes at 72 °C, 30 seconds at 98 °C, 10 cycles of: 98 °C for 10 seconds, 63 °C for 30 seconds, 72 °C for 30 seconds. Final libraries were purified using the MinElute PCR Purification Kit.

Illumina libraries were sequenced in a 2 × 150 bp format using the Novogene sequencing service.

Nanopore sequencing was carried out using Flongle flowcells and the SQK-LSK114 library generation kit following the manufacturer's instructions.

Genome assembly

The `guppy` base caller (version 6.4.6) was used for nanopore base calling with the following settings: `--flowcell FLO-FLG114 --kit SQK-LSK114 --recursive --trim_strategy none`.

Genome assembly was generated using the SPAdes assembler[91] (version 3.15.4) in a hybrid nanopore/Illumina mode.

Genome annotation

Genome annotation was carried out using the RAST web server[92]. Ribosomal RNAs and other ncRNAs were annotated using Infernal[93] (version 1.1.1) and the RFAM database[94] (version 12.0) and RNAmmer[95] (version 1.2).

Sequence analysis

DNA/RNA and protein multiple sequence alignment was carried out using[96].

Ribosomal RNA phylogenetic trees were built using RAxML-NG[4] (version 1.2.0) with the following settings: `--model GTR+G`, with all full-length or near full-length 16S

rRNA sequences for Bdellovibrionota strains available on GenBank.

Protein sequences were translated from the newly generated genome annotation and available genome annotations for other sequenced Bdellovibrionota strains, and then protein domains were annotated using HMMER3[97] (version 3.2.1) and the PFAM database[98] (version 32.0)

Short-read 5-methylcytosine profiling

Genomic DNA was sheared on a Covaris E220, then converted into sequencing libraries following the EM-seq protocol, using the NEBNext Enzymatic Methyl-seq Kit (NEB, Cat # E7120L), and sequenced as 2 × 150mers on a NovaSeq X through Novogene.

Adapters were trimmed from sequencing reads using Trimmomatic[101] (version 0.36). Trimmed reads were aligned against our *de novo* genome assembly using `bwa-meth` with default settings. Duplicate reads were removed using `picard-tools` (version 1.99). Methylation calls were extracted using `MethylDackel` (`https://github.com/dpryan79/MethylDackel`). Additional analyses were carried out using custom-written Python scripts (`https://github.com/georgimarinov/GeorgiScripts`).

ATAC-seq experiments

ATAC-seq experiments were carried out as previously described[31,42].

Cells were fixed by adding 37% formaldehyde (Sigma) at a final concentration of 1% and incubating for 15 minutes at room temperature. Formaldehyde was then quenched using 2.5 M glycine at a final concentration of 0.25 M. Cells were subsequently centrifuged at 13,000 RPM for 1 minute, washed once in 1× PBS, and centrifuged again at 13,000 RPM for 1 minute. Lysis was then carried out by resuspending cells in 400 μL Permeabilization Buffer (33 mM Tris-HCl pH 8.0, 20% sucrose, 25 μg/mL lysozyme) and incubating 5 minutes at 30 °C in a ThermoMixer (with shaking at 1000 RPM). Cells were again pelleted at 13,000 RPM for 1 minute, resuspended in 50 μL transposition mix (25 μL 2× TD buffer, 2.5 μL Tn5, 22.5 μL ultrapure H_2O), and incubated at 37 °C for 15 minutes. The reaction was stopped with the addition of 150 μL IP Elution Buffer (1% SDS, 0.1 M $NaHCO_3$) and 2 μL Proteinase K (Promega, Cat # MC5005), then incubated at 65 °C overnight to reverse crosslinks. DNA was isolated by adding an equal volume of 25:24:1 phenol:chloroform:isoamyl solution, vortexing and centrifuging for 3 minutes at 14,000 rpm, then purifying the top aqueous phase using the MinElute PCR Purification Kit, eluting in 10 μL EB buffer. Libraries were generated as already described above.

ATAC-seq data processing

Demultipexed FASTQ files were mapped to the *de novo* *Bacteriovorax* sp. genome assembly as 2×36mers using Bowtie[99] (version 1.0.1) with the following settings: `-v 2 -k 2 -m 1 --best --strata`. Duplicate reads were removed using `picard-tools` (version 1.99).

TSS scores were calculated as the ratio of ATAC signal in the region ±100 bp around TSSs versus the ATAC signal of the 100-bp regions centered at the two points ±2 kbp of the TSS as previously described[43].

Peak calling was carried out using MACS2[100] with the following settings: `-g 4150000 -f BAM --to-large --keep-dup all --nomodel`.

KAS-seq experiments

KAS-seq experiments were carried out following the previously published protocol[35] with some modifications in the sequencing library generation part[31].

Briefly, a 500-mM N_3-kethoxal solution was brought to 37 °C, then added to 1 mL of culture at a final concentration of 5 μM. Cells were then incubated for 5 minutes at 30 °C in a ThermoMixer at 1000 RPM.

Cells were then pelleted by centrifugation at 13,000 RPM for 1 minute, resuspended in 100 μL 1× PBS buffer, and DNA was immediately isolated using the Monarch Genomic DNA Purification Kit (NEB, Cat # T3010S), with the modification that elution was carried out with 87.5 μL 25 mM K_3BO_3 solution (pH 7.0).

Biotin was clicked onto kethoxal-modified guanines by mixing 87.5 μL DNA, 2.5 μL 20 mM DBCO-PEG4-biotin (Sigma, Cat # 760749; DMSO solution), and 10 μL 10× PBS, and incubating at 37 °C for 90 minutes.

DNA was isolated using AMPure XP beads and eluted in 130 μL 25 mM K_3BO_3 (pH 7.0), then sheared on a Covaris E220 for 120 seconds down to ∼150-200 bp.

Libraries were built on beads using the NEBNext Ultra II DNA Library Prep kit (NEB, Cat # E7645L). Biotin pull down was initiated py pipetting 20 μL Dynabeads My-One Streptavidin T1 beads (ThermoFisher Scientific, Cat # 65306) into DNA lo-bind tubes. Beads were separated on magnet, resuspended in 200 μL of 1× TWB buffer (Tween Washing Buffer; 5 mM Tris-HCl pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20), then separated on magnet again and resuspended in 300 μL of 2× BB (Binding Buffer; 10 mM Tris-HCl pH 7.5, 1 mM EDTA; 2 M NaCl). The DNA (130 μL) was added together with 170 μL 0.1× TE buffer, and incubated at RT on rotator for ≥15 minutes. Beads were separated on magnet, resuspended in 200 μL of 1× TWB, and incubated at 55 °C in a Termomixer for 2 minutes with shaking at 1000 rpm. Beads were again separated on magnet and the 200-μL 55 °C TWB wash step was repeated. Beads were separated on magnet and resuspended in 50 μL 0.1× TE.

End repair was carried out by adding 7 μL NEB End Repair Buffer and 3 μL NEB End Repair Enzyme, incubating at 20 °C for 30 minutes, then at 65 °C for 30 minutes.

End repair was followed by adaptor ligation by adding 2.5 μL NEB Adaptor, 1 μL NEB Ligation Enhancer and 30 μL NEB Ligation Mix, incubating at 20 °C for 20 minutes,

then adding 3 μL USER Enzyme and incubating at 37 °C for 15 minutes. Beads were separated on magnet, resuspended in 200 μL of 1× TWB, then incubated at 55 °C in a Thermomixer for 2 minutes with shaking at 1000 rpm. Subsequently beads were separated on magnet and resuspended in 100 μL of 0.1× TE, separated on magnet again, resuspended in 15 μL of 0.1× TE Buffer, and transfered to PCR tubes.

Beads were then incubated at 98 °C for 10 minutes, and libraries were amplified by adding 5 μL of i5 primer, 5 μL of i7 primer and 25 μL of 2× Q5 Hot Start Polymerase Mix, using the following PCR program: 30 seconds at 98 °C; 15 cycles of 98 °C for 10 seconds, 65 °C for 30 seconds, and 72 °C for 30 seconds; and a final extension at 72 °C for 5 minutes.

Beads were separated on magnet and the final libraries were purified from the supernatant using 50 μL AMPure XP beads, eluting in 0.1× TE buffer.

KAS-seq data processing

Demultipexed FASTQ files were mapped to the *de novo Bacteriovorax* sp. genome assembly as 2×36mers using Bowtie[99] with the following settings: `-v 2 -k 2 -m 1 --best --strata`. Duplicate reads were removed using `picard-tools` (version 1.99).

Differential accessibility/KAS-seq analysis

The analysis of differential chromatin accessibility as measured using ATAC-seq or enriched for KAS-seq signal was carried out using DESeq2[102]. Read counts were calculated over promoters or gene bodies and used as input into DESeq2.

Hi-C experiments

Hi-C experiments were carried out as previously described[87,103] with some modifications.

Cells were crosslinked using 37% formaldehyde (Sigma) at a final concentration of 1% for 15 minutes at room temperature. Formaldehyde was then quenched using 2.5 M Glycine at a final concentration of 0.25 M. Cells were centrifuged at 13,000 rpm for 2 minutes, washed once in 1× PBS, and stored at -80 °C.

Denaturation was carried out by resuspended cells in 50 μL of 0.5% SDS and incubating at 62 °C for 10 minutes. SDS was quenched by adding 145 μL of H_2O and 25 μL of 10% Triton X-100 and incubating at 37 °C for 15 minutes.

Restriction digestion was carried out by adding 25 μL of 10× CutSmart buffer and 100 U of the MluCI restriction enzyme (NEB #R0538) plus 100 U of the MboI enzyme (NEB #R0147), then incubating for ≥2 hours at 37 °C in a Thermomixer at 900 rpm. The reaction was then incubated at 62 °C for 20 minutes in order to inactivate the restriction enzymes.

Fragment ends were filled in by adding 37.5 μL of 0.4 mM biotin-14-dATP (ThermoFisher Scientific, # 19524-016), 1.5 μL each of 10 mM dCTP, dGTP and dTTP, and 8 μL of 5U/μL DNA Polymerase I Large (Klenow) Fragment (NEB #M0210). The reaction was the incubated at 37 °C in a Thermomixer at 900 rpm for 45 minutes.

Fragment end ligation was carried out by adding 663 μL H_2O, 120 μL 10× NEB T4 DNA ligase buffer (NEB B0202), 100 μL of 10% Triton X-100, 12 μL of 10 mg/mL Bovine Serum Albumin (100× BSA, NEB), 5 μL of 400 U/μL T4 DNA Ligase (NEB #M0202), and incubating at room temperature for ≥4 hours with rotation.

Cells were then pelleted by centrifugation at 13,000 rpm for 5 minutes. The pellet was resuspended in 200 μL Elution Buffer (1% SDS, 0.1 M $NaHCO_3$), Proteinase K was added, and incubated at 65 °C overnight to reverse crosslinks.

After addition of 600 μL 1×TE buffer, DNA was sheared using a Covaris E220 instrument. DNA was then purified using the MinElute PCR Purification Kit (Qiagen #28006), with elution in a total volume of 300 μL 1× EB buffer.

For streptavidin pulldown of biotin-labeled DNA, 150 μL of 10 mg/mL Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) were separated on a magnetic stand, then washed with 180 μL of 1× TWB (Tween Washing Buffer; 5 mM Tris-HCl pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20). Beads were then resuspended in 300 μL of 2× Binding Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA; 2 M NaCl), the sonicated DNA was added, and beads were incubated for ≥15 minutes at room temperature on a rotator. After separation on a magnetic stand, the beads were twice washed with 180 μL of 1× TWB, and heated at 55 °C in a Thermomixer with shaking for 2 minutes.

Final libraries were prepared on beads using the NEBNext Ultra II DNA Library Prep Kit (NEB #E7645). End repair was carried out by resuspending beads in 50 μL 1× EB buffer, and adding 3 μL NEB Ultra End Repair Enzyme and 7 μL NEB Ultra End Repair Enzyme, followed by incubation at 20 °C for 30 minutes and then at 65 °C for 30 minutes.

Adapters were ligated to DNA fragments by adding 30 μL Blunt Ligation mix, 1 μL Ligation Enhancer and 2.5 μL NEB Adapter, incubating at 20 °C for 20 minutes, adding 3 μL USER enzyme, and incubating at 37 °C for 15 minutes.

Beads were then separated on a magnetic stand, and washed with 180 μL TWB for 2 minutes at 55 °C at 1000 rpm in a Thermomixer. After separation on a magnetic stand, beads were washed in 100 μL 0.1 × TE buffer, then resuspended in 16 μL 0.1 × TE buffer, and heated at 98 °C for 10 minutes.

For PCR, 5 μL of each of the i5 and i7 NEB Next sequencing adapters were added together with 25 μL 2× NEB Ultra PCR Mater Mix. PCR was carried out with a 98 °C incubation for 30 seconds and 12 cycles of 98 °C for 10 seconds, 65 °C for 30 seconds, and 72 °C for 1 minute, followed by incubation at 72 °C for 5 minutes.

Beads were separated on a magnetic stand, and the supernatant was cleaned up using 1.8× AMPure XP beads.

Libraries were sequenced as 2×150mers on a Illumina NovaSeq X through Novogene.

Hi-C data processing

Hi-C sequencing reads were processed against the *de novo Bacteriovorax* sp. assembly using the Juicer pipeline for analyzing Hi-C datasets[104] (version 1.6 of Juicer and version 2.13.07 of Juicer Tools) with default settings.

Hi-C matrices were visualized using Juicebox[105].

Data availability

The sequencing datasets generated for and used in this study can be accessed from GEO accession XXXX.

Author contributions

G.K.M. conceptualized the study, carried out cell culture, and performed ATAC-seq, KAS-seq and Hi-C experiments together with B.D., analyzed the data and wrote the manuscript, with input from all authors. B.D. performed EM-seq experiments. W.J.G. and A.K. supervised the study.

Acknowledgments

The authors would like to thank members of the Greenleaf and Kundaje labs for helpful discussions. This work was supported by NIH grants (P50HG007735, RO1 HG008140, U19AI057266 and UM1HG009442 to W.J.G., 1UM1HG009436 to W.J.G. and A.K., 1DP2OD022870-01 and 1U01HG009431 to A.K.), the Rita Allen Foundation (to W.J.G.), the Baxter Foundation Faculty Scholar Grant, and the Human Frontiers Science Program grant RGY006S (to W.J.G). W.J.G is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative.

Competing interests

W.J.G. is a consultant and equity holder for 10x Genomics, Guardant Health, Quantapore, and Ultima Genomics, and cofounder of Protillion Biosciences, and is named on patents describing ATAC-seq.

A.K. is a consulting Fellow with Illumina, a member of the SAB of OpenTargets (GSK), PatchBio, SerImmune and a scientific co-founder of RavelBio.

References

1. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**(9):1639–1645.
2. Grigoriev A. 1998. Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26**(10):2286–2290.
3. Lobry JR. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13**(5):660–65
4. 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**(21):4453–4455.
5. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**(7873):583–589.
6. Marinov GK, Lynch M. 2015. Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 (Bethesda)* **6**(2):397–422.
7. Postberg J, Forcob S, Chang WJ, Lipps HJ. 2010. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol* **10**:259.
8. Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**(5532):1074–1080.
9. Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**(11):5088–5090.
10. Lake JA, Henderson E, Oakes M, Clark MW. 1984. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc Natl Acad Sci U S A* **81**(12):3786–3790.
11. Lake JA. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**(6152):184–186.
12. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. 2008. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci U S A* **105**(51):20356–20361.
13. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**(7551):173–179.
14. Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol* **11**(5):209
15. Koonin EV, Yutin N. 2014. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb Perspect Biol* **6**(4):a016188

16. Sandman K, Reeve JN. 2006. Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol* **9**:520–525.
17. Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A* **88**(22):10148–10152
18. Henneman B, van Emmerik C, van Ingen H, Dame RT. 2018. Structure and function of archaeal histones. *PLoS Genet* **14**(9):e1007582.
19. Stevens KM, Swadling JB, Hocher A, Bang C, Gribaldo S, Schmitz RA, Warnecke T. 2020. Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc Natl Acad Sci U S A* **117**(52):33384–33395.
20. Maruyama H, Harwood JC, Moore KM, Paszkiewicz K, Durley SC, Fukushima H, Atomi H, Takeyasu K, Kent NA. 2013. An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Rep* **14**(8):711–717.
21. Mattiroli F, Bhattacharyya S, Dyer PN, White AE, Sandman K, Burkhart BW, Byrne KR, Lee T, Ahn NG, Santangelo TJ, Reeve JN, Luger K. 2017. Structure of histone-based chromatin in Archaea. *Science* **357**(6351):609–612.
22. Bowerman S, Wereszczynski J, Luger K. 2021. Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. *Elife* **10**:e65587
23. Henneman B, Brouwer TB, Erkelens AM, Kuijntjes GJ, van Emmerik C, van der Valk RA, Timmer M, Kirolos NCS, van Ingen H, van Noort J, Dame RT. 2021. Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res* **49**(8):4338–4349.
24. Alva V, Lupas AN. 2019. Histones predate the split between bacteria and archaea. *Bioinformatics* **35**(14):2349–2353.
25. Schwab S, Boyle AL, Dame RT. 2023. Novel histones and histone variant families in prokaryotes. *bioRxiv* 2023.06.01.543357
26. Hocher A, Laursen SP, Radford P, Tyson J, Lambert C, Stevens KM, Picardeau M, Sockett RE, Luger K, Warnecke T. 2023. Histone-organized chromatin in bacteria. *bioRxiv* 2023.01.26.525422.
27. Hu Y, Schwab S, Deiss S, Escudeiro P, Joiner JD, Hartmann MD, Lupas AN, Hernandez Alvarez B, Alva V, Dame RT. 2023. Bacterial histone HBb from Bdellovibrio bacteriovorus compacts DNA by bending. *bioRxiv* 2023.02.26.530074
28. Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader GD, Giaever G, Nislow C. 2012. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* **1**:e00078.
29. Nalabothula N, Xi L, Bhattacharyya S, Widom J, Wang JP, Reeve JN, Santangelo TJ, Fondufe-Mittendorf YN. 2013. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* **14**:391.
30. Badel C, Samson RY, Bell SD. 2022. Chromosome organization affects genome evolution in *Sulfolobus* archaea. *Nat Microbiol* doi: 10.1038/s41564-022-01127-7
31. Marinov GK, Bagdatli ST, Wu T, He C, Kundaje A, Greenleaf WJ. 2022. The chromatin landscape of the euryarchaeon *Haloferax volcanii*. (*bioRxiv* 2022.07.22.501187.)
32. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**(12):1213–1218.
33. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22**(12):2497–2506.
34. Krebs AR, Imanci D, Hoerner L, Gaidatzis D, Burger L, Schübeler D. 2017. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol Cell* **67**(3):411–422.e4.
35. Wu T, Lyu R, You Q, He C. 2020. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity *in situ*. *Nat Methods* **17**(5):515–523.
36. Lieberman–Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**(5950):289–293.
37. Stolp H, Starr MP. 1963. *Bdellovibrio bacteriovorus* gen. et sp. n., a predatory, ectoparasitic, and bacteriolytic microorganism. *Antonie Van Leeuwenhoek* **29**:217–248.
38. Stolp H, Starr MP. 1965. Bacteriolysis. *Annu Rev Microbiol* **19**:79–104.
39. Lovering AL, Sockett RE. 2021. Microbe Profile: *Bdellovibrio bacteriovorus*: a specialized bacterial predator of bacteria. *Microbiology (Reading)* **167**(4):001043
40. Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. *Curr Opin Microbiol* **8**(4):466–472.
41. Oren A, Garrity GM. 2021. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol* **71**(10).
42. Melfi MD, Lasker K, Zhou X, Shapiro L. 2021. ATAC-seq reveals megabase-scale domains of a bacterial nucleoid. *bioRxiv* 2021.01.09.426053

43. Marinov GK, Shipony Z. 2021. Interrogating the Accessible Chromatin Landscape of Eukaryote Genomes Using ATAC-seq. *Methods Mol Biol* **2243**:183–226.
44. Shipony Z, Marinov GK, Swaffer MP, Sinnott-Armstrong NA, Skotheim JM, Kundaje A, Greenleaf WJ. 2020. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat Methods* **17**(3):319–327.
45. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A* **108**:18318–18323
46. Core L, Adelman K. 2019. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**(15–16):960–982.
47. Gilmour DS, Lis JT. 1986. RNA polymerase II interacts with the promoter region of the noninduced *hsp70* gene in *Drosophila melanogaster* cells. *Mol Cell Biol* **6**:3984–3989.
48. Rougvie AE, Lis JT. 1988. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**:795–804.
49. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**(12):1507–1511.
50. Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**:335–338.
51. Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**:950–953.
52. Williams LH, Fromm G, Gokey NG, Henriques T, Muse GW, Burkholder A, Fargo DC, Hu G, Adelman K. 2015. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol Cell* **58**:311–322.
53. Chivu AG, Abuhashem A, Barshad G, Rice EJ, Leger MM, Vill AC, Wong W, Brady R, Smith JJ, Wikramanayake AH, Arenas-Mena C, Brito IL, Ruiz-Trillo I, Hadjantonakis AK, Lis JT, Lewis JJ, Danko CG. 2023. Evolution of promoter-proximal pausing enabled a new layer of transcription control. *bioRxiv* 2023.02.19.529146
54. Kassavetis GA, Chamberlin MJ. 1981. Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *J Biol Chem* **256**:2777–2786.
55. Yanofsky C. 1981. Attenuation in the control of expression of bacterial operons. *Nature* **289**:751–758.
56. Winkler ME, Yanofsky C. 1981. Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. *Biochemistry* **20**:3738–3744.
57. Lau LF, Roberts JW, Wu R. 1983. RNA polymerase pausing and transcript release at the lambda tR1 terminator in vitro. *J Biol Chem* **258**:9391–9397.
58. Kireeva ML, Kashlev M. 2009. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc Natl Acad Sci U S A* **106**(22):8900–8905
59. Kang JY, Mishanina TV, Landick R, Darst SA. 2019. Mechanisms of Transcriptional Pausing in Bacteria. *J Mol Biol* **431**(20):4007–4029.
60. Conconi A, Widmer RM, Koller T, Sogo JM. 1989. Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell* **57**(5):753–761.
61. Merz K, Hondele M, Goetze H, Gmelch K, Stoeckl U, Griesenbeck J. 2008. Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. *Genes Dev* **22**(9):1190–1204.
62. Dwidar M, Im H, Seo JK, Mitchell RJ. 2017. Attack-Phase *Bdellovibrio bacteriovorus* Responses to Extracellular Nutrients Are Analogous to Those Seen During Late Intraperiplasmic Growth. *Microb Ecol* **74**(4):937–946.
63. Price MN, Arkin AP, Alm EJ. 2006. The life-cycle of operons. *PLoS Genet* **2**(6):e96.
64. Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo FY, Pratap A, Deutsch EW, Peterson A, Martin D, Baliga NS. 2009. Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol* **5**:285.
65. Babski J, Haas KA, Näther-Schindler D, Pfeiffer F, Förstner KU, Hammelmann M, Hilker R, Becker A, Sharma CM, Marchfelder A, Soppa J. 2016. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* **17**(1):629
66. Laass S, Monzon VA, Kliemt J, Hammelmann M, Pfeiffer F, Förstner KU, Soppa J. 2019. Characterization of the transcriptome of *Haloferax volcanii*, grown under four different conditions, with mixed RNA-Seq. *PLoS ONE* **14**(4):e0215986
67. Grüberger F, Knüppel R, Jüttner M, Fenk M, Borst A, Reichelt R, Hausner W, Soppa J, Ferreira-Cerca S, Grohmann D. 2019. Exploring prokaryotic transcription, operon structures, rRNA maturation and modifications using Nanopore-based native RNA sequencing. *bioRxiv* 2019.12.18.880849
68. Le TB, Imakaev MV, Mirny LA, Laub MT. 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**(6159):731–734
69. Yildirim A, Feig M. 2018. High-resolution 3D models of *Caulobacter crescentus* chromosome reveal genome structural variability and organization. *Nucleic Acids Res* **46**(8):3937–3952.
70. Trussart M, Yus E, Martinez S, BaÃ¹ D, Tahara

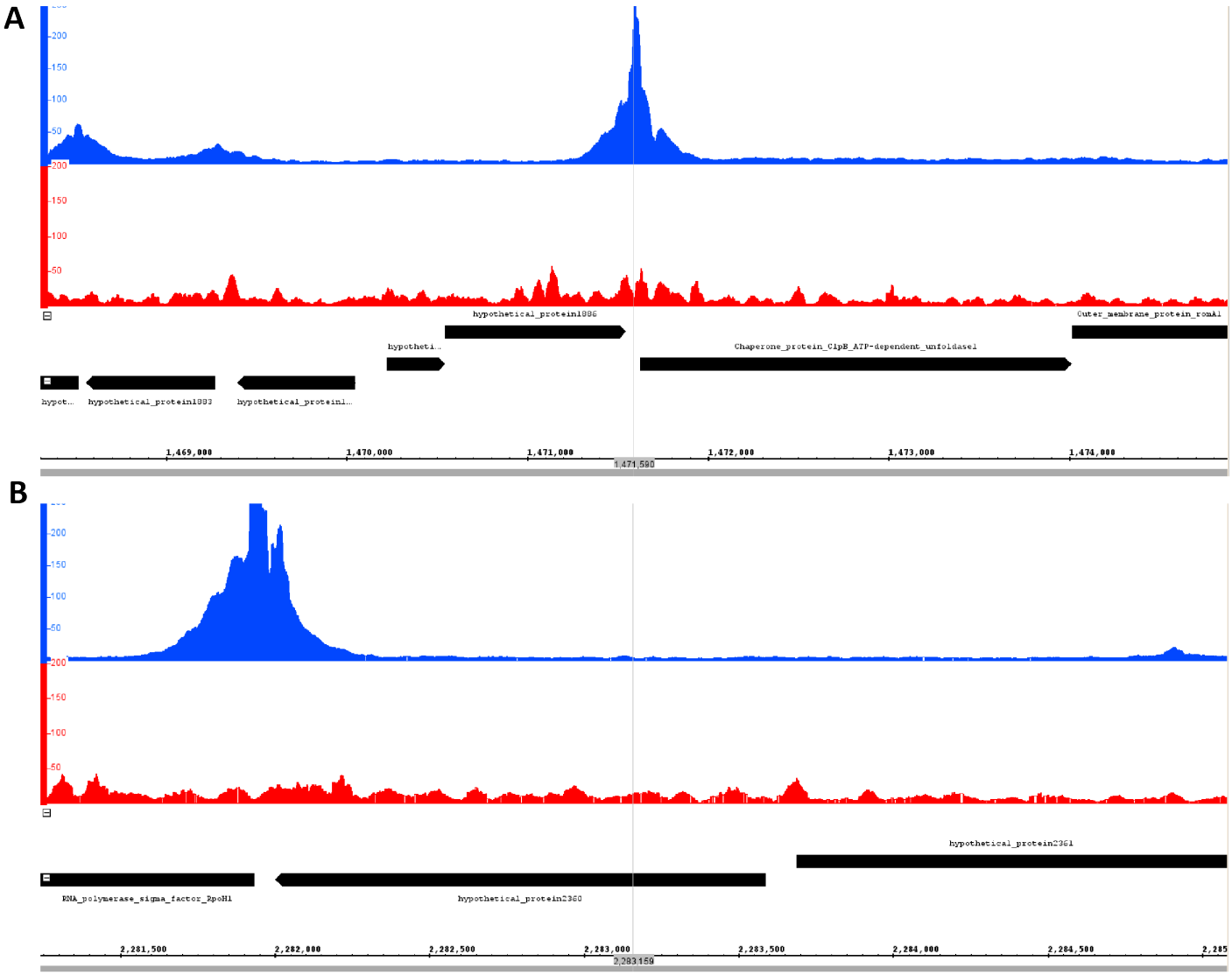
YO, Pengo T, Widjaja M, Kretschmer S, Swoger J, Djordjevic S, Turnbull L, Whitchurch C, Miyata M, Marti-Renom MA, Lluch-Senar M, Serrano L. 2017. Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*. *Nat Commun* **8**:14665

71. Marbouty M, Cournac A, Flot JF, Marie-Nelly H, Mozziconacci J, Koszul R. 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* **3**:e03318
72. Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche JB, Mozziconacci J, Murray H, Koszul R, Nollmann M. 2015. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol Cell* **59**(4):588–602.
73. Marbouty M, Baudry L, Cournac A, Koszul R. 2017. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Sci Adv* **3**(2):e1602105
74. Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espéli O, Boccard F, Koszul R. 2018. Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172**(4):771–783.e18.
75. Böhm K, Giacomelli G, Schmidt A, Imhof A, Koszul R, Marbouty M, Bramkamp M. 2020. Chromosome organization by a conserved condensin-ParB system in the actinobacterium *Corynebacterium glutamicum*. *Nat Commun* **11**(1):1485.
76. Marbouty M, Thierry A, Millot GA, Koszul R. 2021. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. *Elife* **10**:e60608
77. Lamy-Besnier Q, Bignaud A, Garneau JR, Titecat M, Conti DE, Von Strempel A, Monot M, Stecher B, Koszul R, Debarbieux L, Marbouty M. 2023. Chromosome folding and prophage activation reveal specific genomic architecture for intestinal bacteria. *Microbiome* **11**(1):111
78. Wang X, Brandão HB, Le TB, Laub MT, Rudner DZ. 2017. *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science* **355**(6324):524–527.
79. Austin S, Abeles A. 1983. Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. *J Mol Biol* **169**(2):373–87.
80. Austin S, Abeles A. 1983. Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. *J Mol Biol* **169**(2):353–372.
81. Abeles AL, Friedman SA, Austin SJ. 1985. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. *J Mol Biol* **185**(2):261–272.
82. Jalal ASB, Le TBK. 2020. Bacterial chromosome segregation by the ParABS system. *Open Biol* **10**(6):200097
83. Livny J, Yamaichi Y, Waldor MK. 2007. Distribution of centromere-like parS sites in bacteria: insights from comparative genomics. *J Bacteriol* **189**(23):8693–8703.
84. Takemata N, Samson RY, Bell SD. 2019. Physical and Functional Compartmentalization of Archaeal Chromosomes. *Cell* **179**(1):165–179.
85. Cockram C, Thierry A, Gorlas A, Lestini R, Koszul R. 2021. Euryarchaeal genomes are folded into SMC-dependent loops and domains, but lack transcription-mediated compartmentalization. *Mol Cell* **81**(3):459–472.e10
86. Ofer S, Blombach F, Erkelens AM, Barker D, Soloviev Z, Schwab S, Smollett K, Matelska D, Fouqueau T, van der Vis N, Kent NA, Thalassinos K, Dame RT, Werner F. 2023. DNA-bridging by an archaeal histone variant via a unique tetramerisation interface. *Commun Biol* **6**(1):968
87. Marinov GK, Trevino AE, Xiang T, Kundaje A, Grossman AR, Greenleaf WJ. 2021. Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*. *Nat Genet* **53**(5):613–617.
88. Sakrikar S, Schmid AK. 2021. An archaeal histone-like protein regulates gene expression in response to salt stress. *Nucleic Acids Res* **49**(22):12732–12743.
89. Jevtic Z, Stoll B, Pfeiffer F, Sharma K, Urlaub H, Marchfelder A, Lenz C. 2019. The Response of *Haloferax volcanii* to Salt and Temperature Stress: A Proteome Study by Label-Free Mass Spectrometry. *Proteomics* **19**(20):e1800491
90. Dulmage KA, Todor H, Schmid AK. 2015. Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *mBio* **6**(5):e00649–15
91. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**(5):455–477.
92. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**:75.
93. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-

fold faster RNA homology searches. *Bioinformatics* **29**:2933–2935.
94. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD. 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* **43**(Database issue):D130–137.
95. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**(9):3100–3108.
96. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**(5):1792–1797.
97. Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**(10):e1002195.
98. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* **42**(Database issue):D222–230.
99. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3):R25.
100. Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**(9):1728–1740.
101. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120.
102. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**(12):550.
103. Marinov GK, Chen X, Wu T, He C, Grossman AR, Kundaje A, Greenleaf WJ. 2022. The chromatin organization of a chlorarachniophyte nucleomorph genome. *Genome Biol* **23**(1):65
104. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**(1):95–98.
105. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**(1):99–101.

Supplementary Materials

Supplementary Figures



Supplementary Figure 1: Additional examples of extremely strongly paused promoters in *Bacteriovorax*. (A) Chaperone protein ClpB ATP-dependent unfoldase (B) RNA Polymerase Sigma Factor RpoH1.