A giant virus genome is densely packaged by stable nucleosomes

Terri D. Bryson[1,2]†, Pablo De Ioannes[3]†, Marco Igor Valencia-Sánchez[3], Jorja G. Henikoff[1], Paul B. Talbert[1,2], Bernard La Scola[4,5], Karim-Jean Armache[3]*, Steven Henikoff[1,2]*

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center; Seattle, WA, 98109, USA

[2]Howard Hughes Medical Institute; Chevy Chase, MD, USA

[3]Skirball Institute of Biomolecular Medicine, Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY, USA

[4]Institut de Recherche Pour le Développement (IRD), Assistance Publique – Hôpitaux de Marseille (AP-HM), MEPHI, Aix-Marseille University, Marseille, France.

[5]IHU Méditerranée Infection, Marseille, France

†Equal contributions

*Co-corresponding authors. Email: karim-jean.armache@nyulangone.org, steveh@fredhutch.org

Abstract

The doublet histones of Marseillevirus are distantly related to the four eukaryotic core histones and wrap DNA to form remarkably similar nucleosomes. By releasing Marseillevirus chromatin from virions into solution and performing genome-wide nuclease digestion and chemical cleavage assays, we find that the higher-order organization of Marseillevirus chromatin differs greatly from that of eukaryotes. Marseillevirus nucleosomes fully protect DNA within virions, without linker DNA or phasing along genes. Likewise, we observe that most nucleosomes reconstituted onto 3-copy tandem repeats of a nucleosome positioning sequence are tightly packed and fully wrapped. We also document repeat generation and instability during viral passage in amoeboid culture. Dense promiscuous packing of fully wrapped nucleosomes rather than 'beads-on-a-string' with genic punctuation suggests a viral genome protection function for doublet histones.

First isolated in 2007, the giant Marseillevirus encapsulates a 368-kb circular genome encoding 457 proteins (*1*), and related viruses of the family Marseilleviridae have since been discovered infecting *Acanthamoeba* species worldwide (*2*). Marseillevirus genomes encode doublet histones that are paired homologs of the four eukaryote core histones, where Hβ-Hα is homologous to eukaryotic H2B and H2A and Hδ-Hγ is homologous to H4 and H3. Phylogenetic analysis places Hα, Hβ, Hγ, Hδ as sister groups respectively to their H2A, H2B, H3 and H4 eukaryotic counterparts, consistent with divergence within the Marseilleviridae from the last eukaryotic common ancestor for all four histone folds (*3*). We and others have previously used biochemical reconstitution and cryoEM imaging to solve the high-resolution structure of Marseillevirus nucleosomes, which show a striking resemblance to their eukaryotic counterparts (*4, 5*). Unlike octameric eukaryotic nucleosomes, which wrap 147 bp of DNA, tetrameric Marseillevirus nucleosomes wrap only 121 bp of DNA, despite being reconstituted on the Widom 601 artificial positioning sequence, selected for 147-bp wrapping of nucleosome cores.

Because Marseillevirus doublets assemble into nucleosomes that are not fully wrapped by DNA, it was assumed that they are inherently unstable, perhaps to facilitate expression during early stages of infection or for gene regulation (*5*). However, without isolation of viral chromatin in its native form, the extent to which reconstituted Marseillevirus nucleosomes are representative of their conformation in virions is debatable (*6, 7*). By mapping the chromatin landscape of the viral genome we aimed to understand the functional and evolutionary basis for viral nucleosomal packaging.

The capsids of giant viruses that infect amoeba are resistant to treatments that disrupt other viruses, however by dialyzing in low pH conditions Schrad and co-workers (*8*) have shown that giant Mimivirus capsids can be breached, although without releasing their contents. We followed their protocol to open the Marseillevirus capsid, then dialyzed into a neutral buffer for Micrococcal Nuclease (MNase) digestion. High yields of digested chromatin were obtained from samples treated with 0.5% NP-40 over a time-course ranging from 1 minute to 18 hours (**fig. S1**). In contrast, 0.1% digitonin had no effect. As NP-40 permeabilizes membranes by sequestering lipids, whereas digitonin permeabilizes cell membranes by displacing membrane sterols, our results suggest that the chromatin of Marseilleviridae, like that of giant Mimiviridae (*9-12*), is enclosed within an impermeable lipid membrane that would have protected it from low pH conditions and then permeabilized for MNase to gain access.

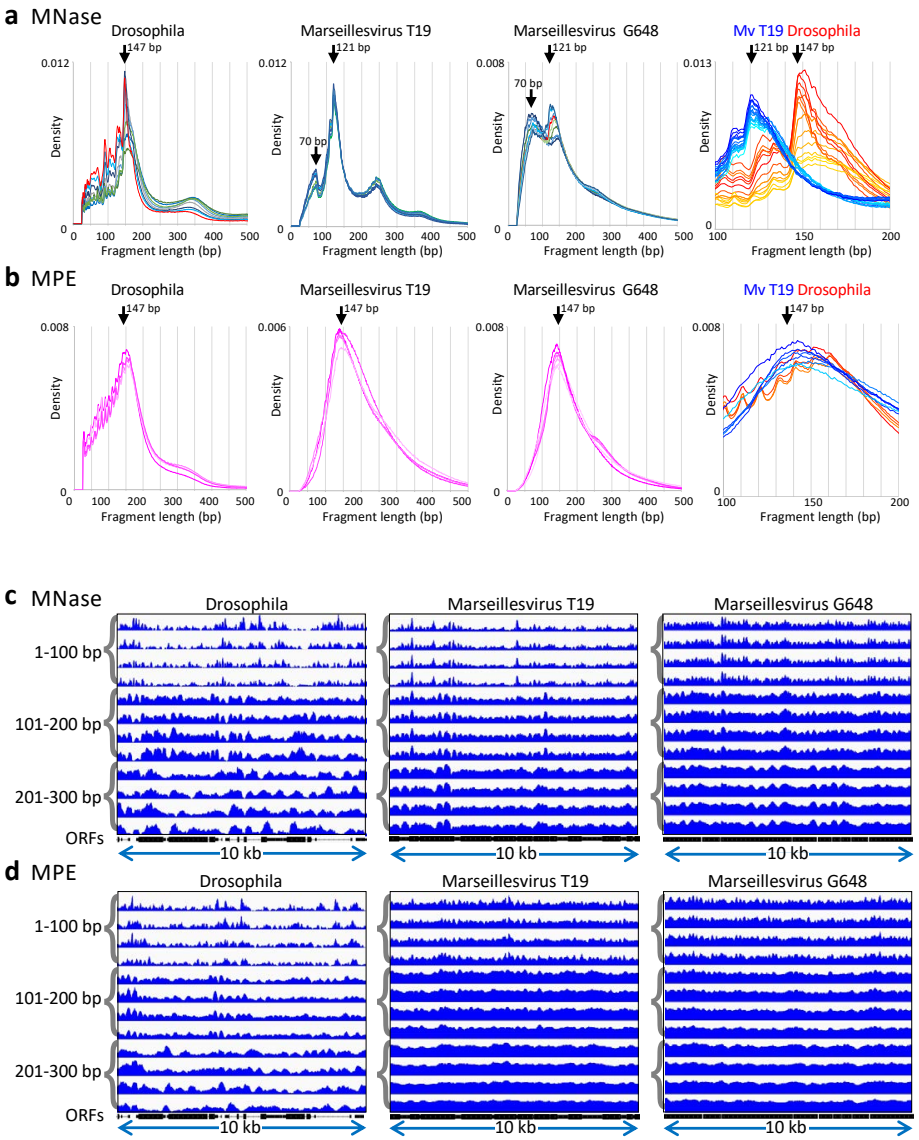

Fig. 1: MNase- and MPE-treated Marseillevirus chromatin yields tightly packed nucleosome-sized DNA fragments. MNase-seq and MPE-seq were performed on Drosophila melanogaster S2 cells and on viral particles after pH 2 treatment and neutralization, followed by Illumina DNA library preparation and PE25 paired-end sequencing. **a**) Drosophila (9 samples), Marseillevirus T19 extracted with NP40 (8 samples) and Marseillevirus G648 (10 samples) extracted without NP40. For MNase-seq, Drosophila and Marseillevirus G648 samples were digested for 5 min in a concentration series starting at 1x (red curve) and decreasing to 1/512x. Marseillevirus T19 samples were extracted for different times in NP40 and digested at the highest (1x) level used for Drosophila and Marseillevirus G648. Graphs on right show a comparison between Drosophila (14 samples) and Marseillevirus T19 (17 samples) on an expanded scale. **b**) Same as (a) for MPE-cleaved samples, Drosophila (4 samples), Marseillevirus T19 (6 samples) and Marseillevirus G648 (4 samples) all extracted with NP40. Curves for individual samples processed in parallel are superimposed to illustrate the degree of variation between replicate samples digested or cleaved under different conditions used in this study. **c-d)** Drosophila and Marseillevirus T19 and G648 MPE-seq and MNase-seq library sequences aligned to their respective genomes, and four MNase samples (c: top to bottom: 6.25, 12.5, 25 and 50 mU/reaction) and four MPE samples (d: top to bottom: 1 µM, 10 µM, 100 µM and 1000 µM H_2O_2) were chosen for comparison. For Drosophila, the 10-kb region around a bidirectional housekeeping promoter pair (His2Av and ball, chr3R:26,862,501-26,872,500) was selected for display. For T19 an arbitrarily chosen representative 10-kb region was selected (53,001-63,000) and the region with the same coordinates in G648 (but not orthologous) was also selected. Fragment size classes represent subnucleosomes (1-100 bp), nucleosomes (101-200 bp) and mixtures of mono- and di-nucleosomes (201-300 bp). Nearly uniform occupancy is seen for MPE-generated Marseillevirus nucleosomes, whereas under the same conditions Drosophila nucleosomes display regions of conspicuous phasing characteristic of nucleosomes separated by linker regions. Tracks are group-autoscaled within sets of four.

Under our permeabilization and digestion conditions, 85-90% of the DNA in virions was recovered as intact MNase-protected particles. At the highest MNase digestion levels, control Drosophila nucleosomes showed a dominant 147-bp peak, but also a smear of sub-nucleosome-sized digestion products (**Fig. 1a**). Using the highest digestion level for Marseillevirus we observed a dominant 121-bp MNase-protected peak, a small ~70-bp peak, but no smear, confirming cryoEM observations of 121-bp wrapping of reconstituted nucleosome cores (*4, 5*). These observations indicate that Marseillevirus nucleosomes are less sensitive to intranucleosomal cleavages than are eukaryotic nucleosomes.

MNase is an endo/exo-nuclease that is known to preferentially digest AT-rich regions (*13, 14*). Consistent with these observations, we find that ~90% of the cleavage sites occur between A/T base pairs (**fig. S2**) and long AT-rich regions are preferentially digested (**fig. S3**). Therefore, we wondered whether the unexpected ~70 bp peak (**Fig. 1a**) resulted from aggressive MNase digestion of AT-rich regions within nucleosomes. To eliminate this potential artifact, we used a small-molecule DNA cleavage reagent, methidiumpropyl-EDTA-Fe(II) (MPE) which hydrolyzes DNA phosphodiester bonds, where H_2O_2 provides reactive oxygen for MPE-Fe(II)-catalyzed DNA cleavage (*15*). MPE-seq is performed similarly to MNase-seq, but without exonuclease activity and without sequence bias (*16*) (**fig. S3**). When we treated Drosophila nuclei and permeabilized Marseillevirus particles with MPE-Fe(II), we observed mostly mononucleosome-sized particles for Drosophila, with a ~150-bp peak and periodic 10-bp internal cleavages (**Fig. 1b**). MPE-seq of Marseillevirus chromatin from both the original isolate (strain T19, later named *Marseillevirus marseillevirus*) and a more recent isolate, G648 (*1*), also revealed a ~150-bp peak with close concordance between samples, and no 70-bp peak. The lack of periodic 10-bp internal cleavages demonstrates that Marseillevirus chromatin is highly refractory to intranucleosomal cleavages (**Fig. 1b**). Whereas MNase digestion resulted in a 26-bp fragment size difference between Marseillevirus and Drosophila nucleosomes (**Fig. 1a** right panel), MPE cleavage resulted in fragments of nearly the same size (**Fig. 1b**, right panel). These consistent discrepancies between MNase- and MPE-generated fragments indicate that the aggressive endo/exonucleolytic activity of MNase and its preference for AT-rich DNA is responsible for the ~70-bp fragment peaks in T19 and G648 (**fig. S2**). Although the genomes of both Marseillevirus isolates have the same ~44% GC-content as Drosophila, they have characteristic A/T homopolymeric runs in promoter regions (*17*) that might contribute to artifactual internal MNase cleavages seen for Marseillevirus nucleosomes.

Interestingly, Marseillevirus nucleosomes remain insoluble even after MNase digestion, in contrast to Drosophila nucleosomes, which are released by MNase into solution (**fig. S4**), as if tight packaging within

the permeabilized virion prevents release of mono-nucleosomes. To examine the global organization of Marseillevirus nucleosomes, we separated fragments by size in 100-bp intervals and displayed a representative 10-kb region of the T19 and G648 genomes for MNase- and MPE-generated replicate samples. Typical Drosophila chromatin profiles showed variable nucleosome positioning with patches of positioned nucleosomes for 101-200 bp fragments using either MPE cleavage or MNase digestion, but not for subnucleosomal 1-100 bp or supranucleosomal 201-300 bp fragments (**Fig. 1c-d**). In contrast, the 101-200 bp MPE-generated fragments from both T19 and G648 nucleosomes showed at most minor variations that were consistent between replicates over a uniform chromatin landscape with no conspicuous positioning in the 101-201 bp subset (**Fig. 1d**). This lack of positioning in Marseillevirus T19 and G648 nucleosomes suggests that they are tightly packed without intervening linkers characteristic of eukaryotic nucleosomes.

We wondered whether the minor regularities seen in the MNase-digested Marseillevirus nucleosome landscapes correspond to genic regions, which are annotated as Open Reading Frames (ORFs) and are separated by very short spans that are usually AT-rich (*17*). To investigate this possibility, we aligned all 191 Open Reading Frames (ORFs) that are ≥600 bp at their 5' ends and averaged each nucleotide position over gene bodies. For comparison we chose the first 191 ORFs ≥600 bp from the chronological list of Drosophila ORFs. Alignment of MNase-generated 101-200 bp fragments to the 191 Drosophila ORFs revealed the characteristic translational phasing pattern, with a prominent +1 nucleosomal peak just downstream of the ORF 5' end and phased +2 and +3 nucleosomes with reduced occupancy for all digestion levels over a 128-fold range (**Fig. 2a**). MPE-generated fragments showed the same 5'-aligned translational phasing patterns for 101-200 bp fragments, confirming that maps produced using MPE and MNase are approximately concordant for control eukaryotic nucleosomes. In contrast, MNase-generated Marseillevirus T19 and G648 101-200 bp fragments displayed very little if any 5' phasing when plotted on the same scale (**Fig. 2a, c**), although when the scale was expanded a phased +1 nucleosome was observed. However, no phasing was seen for MPE-generated T19 and G648 5'-aligned ORFs, indicating that the minor phasing that was seen with MNase is attributable to internal cleavages within nucleosomes. The lack of 5' phasing, a universal characteristic of eukaryotic genes, indicates that the packaged Marseillevirus genome is inactive. Alignment of the 191 ORFs at their 3' ends shows extreme sensitivity of Drosophila nucleosomes to MNase levels not seen for MPE-generated fragments (**Fig. 2b**), which is consistent with partial unwrapping of AT-rich ORF 3'-end DNA from nucleosome cores and sensitivity to MNase exonucleolytic activity. A similar MNase sensitivity and MPE insensitivity is seen for T19 and G648

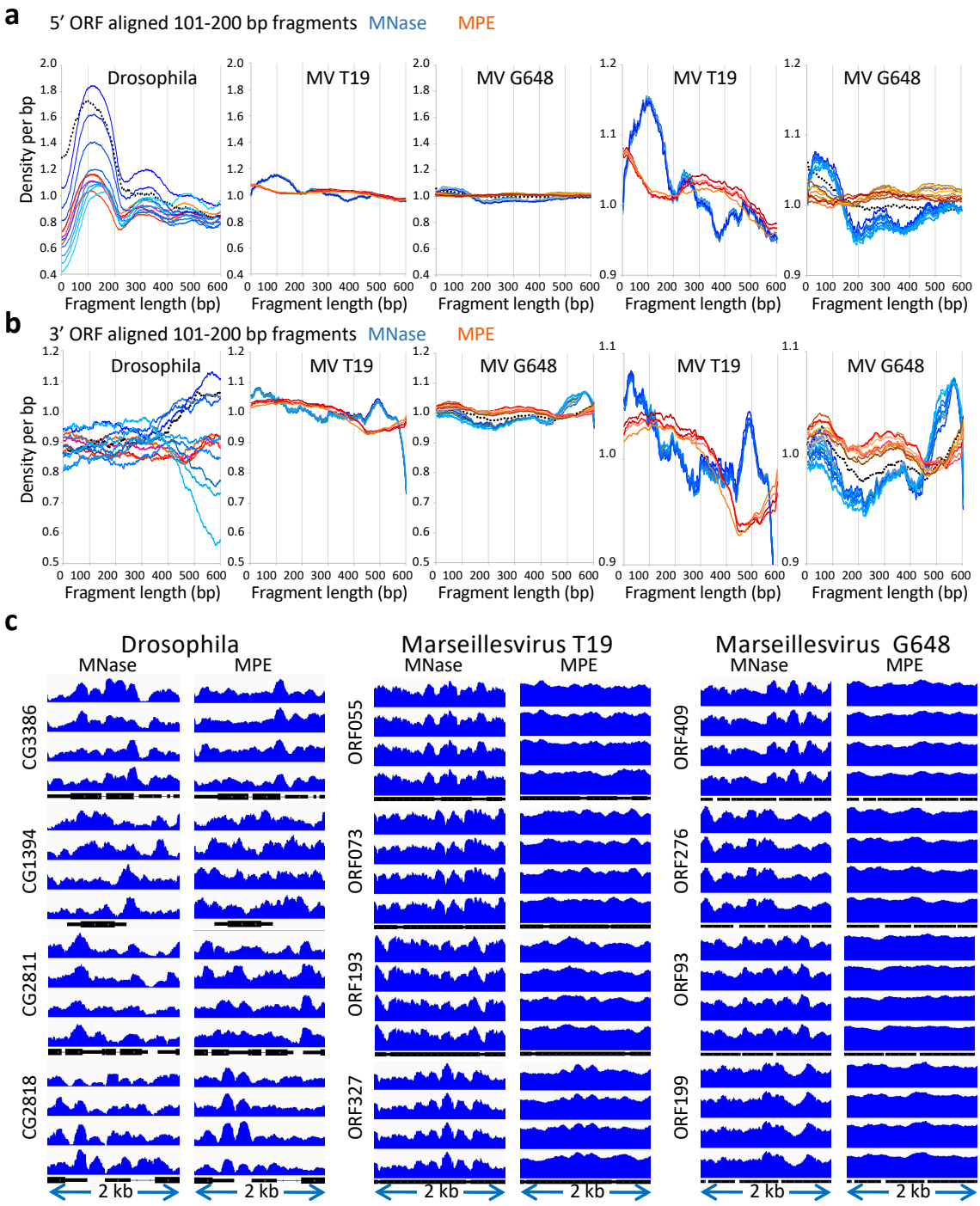

Fig. 2: Phasing observed by MNase-seq for Marseillevirus is not observed by MPE-seq. All 191 Marseillevirus ORFs ≥600 bp were aligned at their 5' ends (**a**) and 3' ends (**b**) and normalized counts were plotted at 1-bp resolution over the 600 bp span. Curves with bluish colors are from MNase-seq for different samples digested in parallel, and curves with reddish colors are for MPE-seq samples. Left to right: Drosophila S2 cells (8 MNase and 4 MPE samples), Marseillevirus T19 (8 MNase and 4 MPE samples) and G648 (10 MNase and 8 MPE samples) plotted on the same scale, and T19 and G648 plotted on a vertically expanded scale to display the differences between MPE and MNase digestions. Dotted line is for a no MNase control. **c**) ORFs showing high variation in periodicity based on autocorrelation (**fig. S5**) downstream of the 5' end for the 101-200 kb size class were displayed within 2-kb regions for four technical replicates each. For MPE-seq, samples were digested using (top to bottom) 1 mM, 10 mM, 100 mM and 1000 mM H_2O_2. For MNase-seq, samples were digested in (top to bottom) 50, 25, 12.5, 6.25 mU MNase/reaction. In contrast to Drosophila nucleosomes, which show phasing of 101-200 bp fragments resulting from either MNase or MPE digestion, Marseillevirus nucleosomes are not detectably phased when assayed by MPE-seq, and MNase-seq shows phasing irrespective of fragment size, indicating internal cleavages. Tracks are group-autoscaled within sets of four.

nucleosomes at the very 3' ends of ORFs (**Fig. 2b**), which is likewise attributable to AT-rich regions at Marseillevirus 3' ends. However, unlike chromatin at the 3' ends of Drosophila ORFs, both T19 and G648 chromatin displayed an average peak of MNase resistance just upstream of the 3' end. As this peak was absent from MPE-digested average profiles, we attribute it to internal MNase cleavage of a subset of nucleosomes that are relatively excluded from neighboring AT-rich regions rather than to 3' nucleosome phasing. The lack of any chromatin accessibility features punctuating genic regions implies that Marseillesvirus nucleosomes have evolved exclusively for packaging within the virion.

Mapping of Marseillevirus fragments to the T19 genome assembly revealed an unexpected over-representation of fragments relative to the rest of the genome in two sharply defined regions (**Fig. 3a**, left panel). A ~20-kb region centered over Position 27,000 is ~3-fold over-represented and an ~2-kb region centered over Position 318,400 is ~10-fold over-represented similarly in all ten MNase datasets regardless of digestion level and in all four MPE datasets regardless of H_2O_2 concentration. To determine whether these striking over-representation features were present in the initial sample from 2007 used to assemble the original T19 map, we mapped the primary T19 reads from the original fastq files, and observed that over-representation at Position 27,000 was already conspicuously present in this sample (**Fig. 3a**, top track). To quantify over-representation, we randomly sampled fragments of similar lengths to each over-represented feature from the rest of the genome 1000 times and plotted the abundance of each sample on a log-log cumulative plot for the raw reads from the original 2007 virus culture and the No-MNase paired-end reads from the 2020 virus culture used in this study (**Fig. 3b**, left panel). This revealed that the 20-kb region over Position 27,000 was ~2-fold over-represented in the original 2007 viral culture and ~2.7-fold over-represented in the 2020 culture relative to their respective genome-wide median abundances. Likewise, the 2-kb region over Position 318,400 was ~1.1-fold over-represented in the original 2007 viral culture and ~10-fold over-represented in the 2020 culture. Although the magnitude of over-representation of the 2-kb region in the original 2007 viral culture is small, it is ranked between #1 and #2 of the 1000 randomly sampled regions, and so likely represents incipient over-representation of some virions in that culture. It is evident that over-representation of the 20-kb region continued to be maintained at 2-3-fold over-abundance during successive passages in *Acanthamoeba polyphaga* culture at similar levels within the cultured Marseillevirus populations, while the 2-kb region expanded ~12-fold during successive passages in culture.

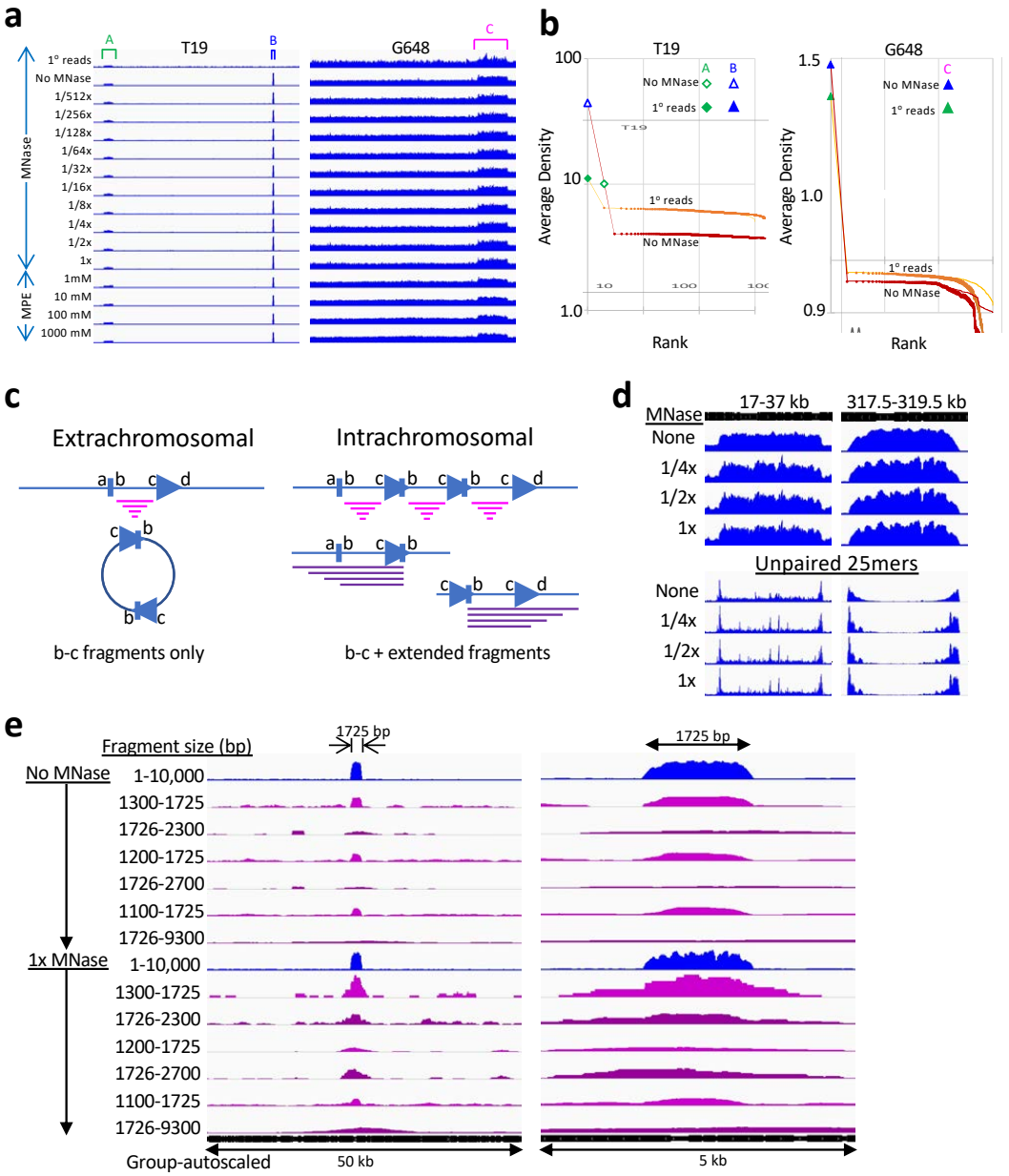

Fig. 3: Regional over-representation of DNA during culturing of Marseillevirus. a) IGV screenshots reveal over-representation of specific regions detected by MNase-seq and MPE-seq and of primary reads used in the T19 and G648 assemblies (*1*). **b**) Cumulative plot showing the rank of each over-represented region among random samples of the span from regions of the rest of the genome, where Rank 1 is the most over-represented sample. We excluded 1 kb on either side of the regions and at either end of the genome assembly. Plots show samples from primary T19 and G648 reads and from No MNase samples in this study. **c**) Two models for regional DNA over-representation of DNA fragments in T19 and G648 samples. Both extrachromosomal and intrachromosomal tandem repeat models predict novel c/b junctions and mappable fragments (magenta) above background that are shorter than the over-represented region, but only the intrachromosomal model predicts extended fragments above background that start upstream of point a and span downstream beyond point b and start downstream of point d and span upstream beyond point c. **d**) To identify novel junctions at the borders of over-represented regions we compare maps of total fragment occupancy to maps of 25-mers where only a single 25-mer fragment end was mapped (Unpaired). Y-axis is autoscaled within groups. A novel c/b junction will be recognized because any fragment spanning the junction without overlap will not be mapped as the 25-mer ends will be oriented 3'-to-5' relative to one another, rather than 5'-to-3'. **e**) Tracks show the region around the 1725-bp over-represented region observed in No-MNase- and MNase-seq data for the indicated fragment size intervals. The enrichment of 1725-bp fragments spanning point b on the left and point c on the right relative to background is seen for all comparisons, as predicted for the intrachromosomal model, but not the extrachromosomal model.

We next asked if regional over-representation and expansion during culturing is a general feature of the Marseilleviridae (*1*). G648 is a much more recent isolate and therefore has been maintained for a shorter time in culture than T19. For G648 we observed a different region of over-representation, in which a ~50 kb segment approximately centered over Position 319,000 is present in all ten samples in the MNase concentration series and identical results were obtained using MPE digestion (**Fig. 3a**, right panel). Although the primary reads from the fastq files used for the original assembly of G648 were relatively sparse, quantitative sampling analysis showed that this ~50-kb region also strongly over-represented, with a 1.9-fold excess, compared to a 2.6-fold excess for the No MNase sample (**Fig. 3b** right panel). We conclude that sporadic regional over-representation is a general feature of Marseillevirus culturing in *A. polyphaga*, even for newly isolated virions. Analysis of long repeat-spanning fragments and unpaired 25mer fragment ends indicated that regional over-representation is accounted for by intrachromosomal tandem repeats rather than extrachromosomal circles (**Fig. 3c-e**).

We wondered whether previous reconstitutions of Marseillevirus nucleosomes on single 601 sequences (*4, 5*) failed to show fuller wrapping because of the lack of neighboring nucleosomes, which are closely abutted in virions. Accordingly, we performed MPE and MNase digestion on native and cross-linked reconstituted Marseillevirus chromatin that had been assembled at 60 µg/ml or 300 µg/ml concentrations onto a three-copy 601 array and onto a 12-copy 601 array. Following DNA extraction, we prepared sequencing libraries and performed paired-end sequencing, aligning fragments from native or cross-linked 3-copy and 12-copy chromatin digests to the 3-copy 601 array. For the 3-copy native chromatin, both MPE and MNase showed similar chromatin landscapes for the 60 µg/ml and 300 µg/ml samples, with higher occupancy over the three 147-bp 601 positioning sequences than over the intervening 40-bp linkers (**Fig. 4a**). Although the net occupancy over positioning sequence and linker differed by less than half, consistent with high occupancy throughout the array, the transitions between 601 and linker sequence were sharply defined, indicating an excess of 147-bp 601 particles precisely positioned over the 601 sequence. A much smaller excess of 147-bp 601 particles was observed for MPE digestion of cross-linked 3-copy chromatin.

To quantify the relative abundance of precise 147-bp 601 particles, we plotted the length distributions for each 3-copy sample (**Fig. 4b**). In each case, MPE digestions of 3-copy array chromatin resulted in 1-bp wide fragment length peaks at 147 bp, 334 bp and 521 bp. As the entire 3-copy array is 521 bp, and 334 bp is precisely the size expected for a 601-linker-601 spanning fragment, it is evident that MPE endonucleolytically digests to completion without detectable encroachment into the nucleosome-wrapped

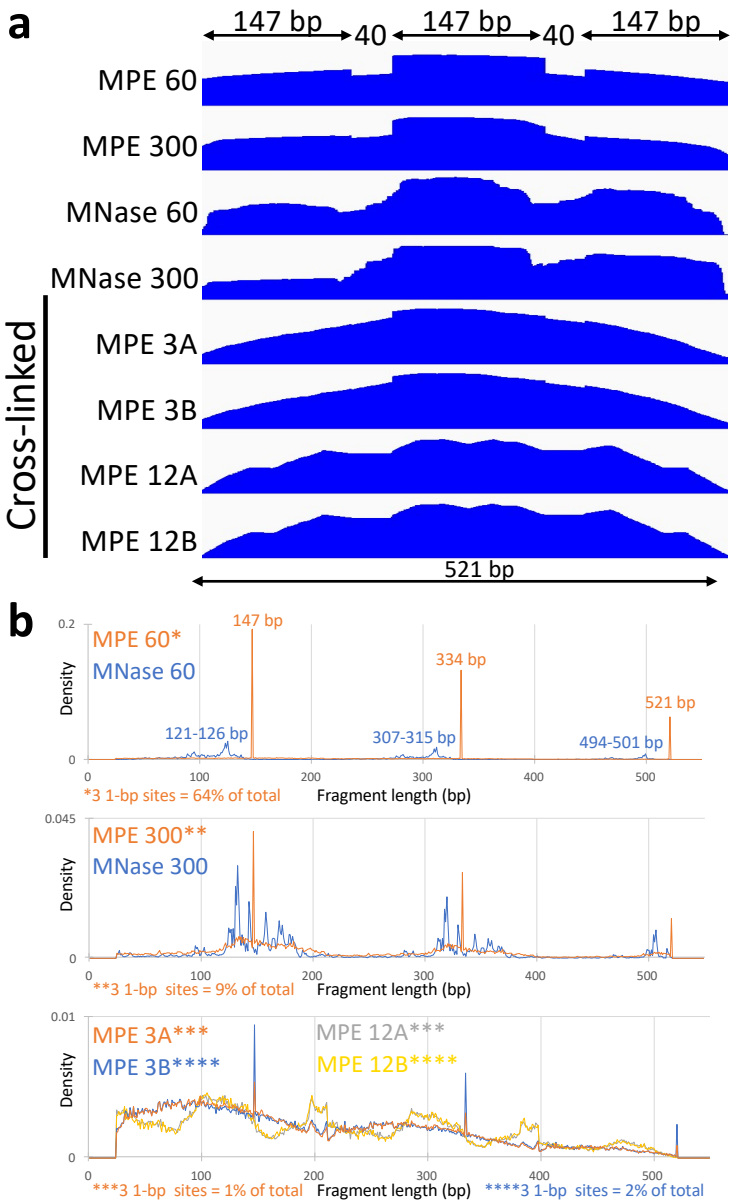

Fig. 4: Assembly of Marseillevirus histones onto Widom 601 arrays produces 147-bp nucleosomes. Three-copy or 12-copy Widom 601 arrays and 60 µg/ml or 300 µg/ml of Marseillevirus histones were assembled into nucleosomes and subjected to MPE-seq and MNase-seq. **a)** Profiles across the 521-bp 3-copy array map were group autoscaled. The higher abundance of fragments towards the middle of each array is a consequence of mapping to a linear fragment, where fragments that span an end can only align with internal copies. MPE3A and MPE3B were assembled onto a 521-bp 3-copy array and MPE12A and MPE12B were assembled onto a 2084-bp 12-copy Widom 601 array, cross-linked, digested with MPE and aligned to the 3-copy array map. **b)** Length distributions of fragments produced from the 3-copy array reconstituted samples show that ~40% of the total occupancy is accounted for by precise cleavages on either or both ends of a 147-bp particle.

particle. Many particles in the 60 µg/ml sample were precisely phased over the 601 positioning sequence, as 147 bp, 334 bp and 521 bp fragments accounted for ~40% of the total assembled chromatin. One basepair wide peaks at these positions in the plot were also seen for cross-linked 3-copy Marseillevirus chromatin arrays, superimposed over a broad distribution of fragment lengths indicative of densely packed chromatin. In contrast to the results with MPE, MNase digestion of the same reconstituted chromatin

showed a distribution of fragment lengths ~20-25 bp smaller than the three discrete fragment lengths produced by MPE digestion. This difference in fragment length for reconstituted Marseillevirus chromatin is similar to the difference observed for digestion of viral chromatin, and provides confirmation that assembly onto a tandem array of positioning sequences as opposed to assembly onto a single positioning sequence recapitulates the situation *in virio*. Cross-linking resulted in reduced precision of MPE digestion for 3-copy array chromatin and total absence of 1-bp peaks for the cross-linked 12-copy array, which suggests that inter-nucleosomal cross-links of closely abutted nucleosomes interfered with the penetration or cleavage activity of MPE.

In summary, we have shown that Marseillevirus nucleosomes can be recovered intact from virions and used to elucidate nucleosome organizational features and document genomic instability using both MNase-seq and MPE-seq. These methods reveal particles that differ from eukaryotic nucleosomes in being refractory to internal cleavages and tightly packed into a landscape without linker DNA or phasing around genes. MPE-seq also shows that the nucleosomes of two Marseillevirus isolates protect ~147 bp of DNA, confirmed by the observation of fully wrapped 147-bp reconstituted particles on Widom 601 DNA. Taken together, our findings reveal a dense chromatin landscape that may have evolved to maximize protection of viral DNA for survival during infection in amoeba cytoplasm. Following infection, the tightly packed undifferentiated Marseillevirus chromatinized genome, surrounded by a lipid membrane (*9-12*), becomes incorporated into a "viral factory" where it undergoes cycles of replication and gene expression within *Acanthamoeba* cytoplasm, cytologically separate from the cell nucleus (*5, 12*). This separation may have facilitated viral histones specializing for chromatin condensation rather than gene regulation.

Materials and Methods

Viruses

Marseillevirus marseillevirus Strains T19 (*2*) and G648 (*1*) were cultured in *Acanthamoeba polyphaga* as described (*18*).

Drosophila cells

Drosophila S2 cells were grown in HyClone SFX Insect Cell Culture Media (Cytiva SH30278.02) supplemented with 18mM L-Glutamine, seeded at $2x10^6$/mL three times per week, and harvested with >95% viability at mid-log phase. A total of $1x10^7$ cells were centrifuged in a swinging-bucket rotor for 4 min at 700xg at 25°C and washed twice in cold 1x PBS. The cell pellet was resuspended in 1 mL TM2+PI

(10mM Tris pH 8, 2mM $MgCl_2$ + Protease Inhibitor, Sigma 11836170001) and chilled in ice water for 1 min. NP-40 was added to 0.5% and vortexed gently at half maximum speed for ~3 sec and returned to ice water slurry. Release of nuclei was ascertained by microscopic observation of aliquots until ≥80% of cellular membranes were disrupted (~3 min). Nuclei were centrifuged 10 min at 150xg at 4˚C, washed twice in 1.5 mL TM2+PI and finally resuspended in 200 µL TM2+PI. Each digestion reaction contained either 50K or 150K nuclei per timepoint.

Viral capsid opening and permeabilization

We followed the viral opening procedure described by Schrad et al., 2020 (*8*) for Mimivirus, with minor modifications. Each sample from a purified Marseillevirus culture was applied to the membrane of a 7K MWCO Dialysis Unit (Thermo 69562) and dialyzed against 250 mL of 20 mM sodium phosphate buffer adjusted to pH2, 2 mM $MgCl_2$, 1 mM PMSF for 2 hr at 25°C, followed by dialysis against 250 mL TM2, 1 mM PMSF for 3 hr. In some experiments viral suspensions were diluted with TM2 prior to dialysis. To equalize the amounts of DNA from Marseillevirus and Drosophila control cells, DNA was extracted from a range of volumes of purified Marseillevirus culture in parallel with a known number of S2 cells. Either 0.5 µL or 1.5 µL of Marseillevirus culture was used per timepoint.

To improve recovery, we included the non-ionic detergent NP-40, which is widely used for chromatin release from cells. Recovery from Marseillevirus T19 particles was ~5-6%, but increased to ~85-90% in 0.5% NP-40 using the maximum MNase digestion conditions that had resulted in mostly mononucleosomes in the previous Drosophila experiments (**fig. S1c-e**).

Nucleosome reconstitution

Expression and purification of Marseillevirus histones, purification of Widom 601 DNA arrays and nucleosome assembly was performed as previously described (*4*).

MNase-seq

MNase-seq was performed as previously described (*19*). Briefly, Micrococcal Nuclease (MNase, Sigma N3755) was reconstituted to a concentration of 1 unit per 5 µl with nuclease-free water, aliquoted and stored at -20°C. MNase was thawed on ice and diluted stepwise 1:1 from 1x to 1/512x with TM2 buffer (10 mM Tris pH 8, 2 mM $MgCl_2$) where 1x is 20 mU/µL. The volume of each Drosophila or Marseillevirus timepoint was adjusted to 166 µL with TM2+PI (Protease Inhibitor, Sigma 11836170001), and 2.5 µL of each MNase dilution was added with mixing, then heated to 37°C for 1min. MNase was activated by

addition of 3.5 µL of 100 mM $CaCl_2$ and incubated for 5 min at 37°C. Reactions were stopped with 172 µL of 2xSTOP solution (10mM Tris, 2mM $MgCl_2$, 340mM NaCl, 20mM EDTA, 4mM EGTA, 100ug/mL RNase, DNase-free (Sigma11119915001) and DNA was extracted as described below. For reconstituted nucleosomes, we digested with 0.2 U MNase/µg DNA for 7 min at room temperature (*20*).

We performed digestions over a time course and concentration range that we had previously found to be sufficient for digesting chromatin from Drosophila S2 cells into mono- and oligo-nucleosomes (*19*). In that study, 1 minute digestions at 2.5 U/million cells had yielded an electrophoretic 'ladder' dominated by oligonucleosomes, 5 minute digestions yielded mostly mononucleosomes, and 30 minute digestions resulted in partially degraded mononucleosomes (30 min). By Tapestation analysis we observed a Marseillevirus nucleosomal ladder, with slightly smaller mononucleosome fragment sizes and much shorter distances between successive peaks than observed for Drosophila digested under same conditions (**fig. S2b**).

We prepared MNase-seq Illumina sequencing libraries from the resulting MNase-digested fragments for both NP-40-treated and untreated libraries, performed paired-end DNA sequencing, and mapped the resulting read pairs to the annotated Marseillevirus T19 genome assembly. On average 97% of fragments mapped to this assembly, confirming the purity of our viral sample. Fragment lengths inferred from sequencing data are smaller than what we observed by Tapestation analysis of purified DNA following MNase digestion (**fig. S1d**), which reflects the selection for smaller fragments during PCR.

MPE-seq

MPE-seq was performed as described by Ishii et al. (*16*). Briefly, the opened Marseillevirus and S2 nuclei were treated with hydrogen peroxide across the range of 0.001-1 mM and cleavage was induced by addition of methidiumpropyl-EDTA-Fe(II) (MPE, a generous gift from Jim Kadonaga), which is MPE complexed with ammonium iron(II) sulfate) at 10 µM or 40 µM for 5 min at room temperature. The reaction was quenched with 6 mM bathophenanthroline (Sigma 133159), followed by 2xSTOP buffer (10 mM Tris, 2 mM $MgCl_2$, 340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 100 µg/mL RNase, DNase-free) at a volume equal to the sample and total DNA extracted as described below. For reconstituted nucleosomes, we digested with 40 µM MPE using at a ratio of 1 mM H_2O_2/650 ng DNA for 5 min at room temperature.

DNA extraction

Sample volumes were adjusted to ~340 µL with TM2. To each sample, 3.4 µL 10% SDS and 2.5 µL Proteinase K (20 mg/ml) were added and incubated 30 min 55ºC. DNA was extracted once with 350 µL Phenol:Chloroform:Isoamyl Alcohol (25:24:1) in a phase-lock tube (Qiagen cat. no. 129046) and centrifuged 5 min at 16,000xg followed by extraction with 350 µL chloroform. The aqueous layer was transferred to a fresh tube containing 2 µL Glycogen (20 mg/mL Sigma cat. no. 10901393001). DNA was precipitated with three volumes of 100% ethanol and centrifuged at 16,000xg for 30 min at 4˚C.. The pellet was rinsed twice with 1 ml 80% ethanol, air dried and dissolved in 30 uL 10 mM Tris pH 8.0. Two µL was analyzed using a HDS1000 ScreenTape on an Agilent 4200 TapeStation.

Library Preparation

Sequencing libraries were prepared from DNA fragments using the KAPA HyperPrep Kit (KAPA cat. no. KK8504) following the manufacturer's instructions. Libraries were amplified for eight cycles using a 10 sec 60˚C combined annealing/extension step. Alternatively, a 72˚C 1 min PCR extension step was added, with no apparent difference in the length distributions. To deplete total DNA samples of large fragments originating from insoluble chromatin prior to library preparation, samples were mixed with ½ volume of HighPrep™ PCR Clean-up System beads (MagBio Genomics cat. no. AC-60005), held 5–10 min, placed on a magnet stand, and the supernatant was collected. Based on the original sample volume, 0.8 volume of beads was added to the supernatant and held 5–10 min. Tubes were placed on a magnet stand, the supernatant discarded and the beads washed 1x with 80% ethanol prior to eluting in 15 uL 10 mM Tris-HCl pH 8. Library size distributions were resolved on an Agilent 4200 TapeStation.

Data processing and analysis

Barcoded libraries were mixed to achieve equimolar representation as desired aiming for a final concentration as recommended by the manufacturer for sequencing on an Illumina HiSeq 2500 2-lane Turbo flow cell. For each sample the following analysis steps were performed:

1) Aligned Illumina fastq files to a reference genome with bowtie2 2.4.2 using parameters --end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700

a) T19: GCF_000887095.1_ViralProj43573_genomic.fna from NCBI Marseillevirus marseillevirus, strain T19 taxon 1559367

b) G648: G648_26-12-14_Genome_vD.fasta (Supplementary Information)

c) Widom 601 3x Array: Genome_601x3.fasta (Supplementary Information)

d) *Drosophila melanogaster*: dm6 from UCSC. Chronological list of *D. melanogaster* ORFs: https://hgdownload.cse.ucsc.edu/goldenPath/dm6/database/refFlat.txt.gz

2) Extracted aligned fragments from bowtie2 sam files into bed files and divided into four fragment length subgroups: 1-, 1-100, 101-200, 201-300. Steps 3-8 were done for each of the four fragment lengths subgroups.

3) Computed percent GC for 10 base pairs on either end of the mapped fragments.

4) Made normalized count bedgraph files using bedtools 2.30.0 genomecov. Normalized counts are the fraction of counts at each base pair scaled by the size of the reference sequence so that if the scaled counts were uniformly distributed there would be 1 at each position.

5) Obtained annotations and selected ≥ 600 bps long

a) T19: from NCBI: https://www.ncbi.nlm.nih.gov/nuccore/284504040

b) G648: Reference (*1*)

c) *Drosophila melanogaster*: dm6 annotation from UCSC: https://genome.ucsc.edu

6) Extracted normalized count (bedgraph) values for 600 bps inside the selected genes from the TSS and from the TES using deepTools 3.3.1 computeMatrix with these parameters:

-S <bigwig version of bedgraph file> -R genes.600 --referencePoint TSS -b 0 -a 600

-S <bigwig version of bedgraph file> -R genes.600 --referencePoint TES -b 600 -a 0

7) Computed autocorrelation for of the extracted values for 1-300 lags using a custom program as R_lag = 1/((n-lag)*var) * sum_1,n-lag(X_t - mean)(X_t+lag - mean) for each gene in each TSS and TES set.

8) Computed the mean and standard deviation of all 299 lags for each gene and reverse ranked the genes by standard deviation, selecting the four genes with the highest standard deviation for each TSS and TES set.

References

1. D. Sahmi-Bounsiar *et al.*, Marseilleviruses: An Update in 2021. *Front Microbiol* **12**, 648731 (2021).
2. M. Boyer *et al.*, Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21848 (2009).
3. A. J. Erives, Phylogenetic analysis of the core histone doublet and DNA topo II genes of Marseilleviridae: evidence of proto-eukaryotic provenance. *Epigenetics Chromatin* **10**, 55 (2017).
4. M. I. Valencia-Sanchez *et al.*, The structure of a virus-encoded nucleosome. *Nat Struct Mol Biol* **28**, 413 (2021).
5. Y. Liu *et al.*, Virus-encoded histone doublets are essential and form nucleosome-like structures. *Cell* **184**, 4237 (2021).
6. A. Vannini, I. Marazzi, A small nucleosome from a weird virus with a fat genome. *Mol. Cell* **81**, 3447 (2021).
7. Y. Liu, M. Krupovic, Genome chromatinization in giant double-stranded DNA viruses. *Trends Biochem. Sci.* **47**, 3 (2022).
8. J. R. Schrad, J. S. Abrahao, J. R. Cortines, K. N. Parent, Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. *Cell* **181**, 1046 (2020).
9. Y. G. Kuznetsov *et al.*, Atomic force microscopy investigation of the giant mimivirus. *Virology* **404**, 127 (2010).
10. C. Xiao *et al.*, Structural studies of the giant mimivirus. *PLoS Biol* **7**, e92 (2009).

11. N. Zauberman *et al.*, Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS Biol* **6**, e114 (2008).
12. M. Suzan-Monti, B. La Scola, L. Barrassi, L. Espinosa, D. Raoult, Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. *PLoS One* **2**, e328 (2007).
13. H. R. Chung *et al.*, The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One* **5**, e15754 (2010).
14. J. D. McGhee, G. Felsenfeld, Another potential artifact in the study of nucleosome phasing by chromatin digestion with micrococcal nuclease. *Cell* **32**, 1205 (1983).
15. I. L. Cartwright, R. P. Hertzberg, P. B. Dervan, S. C. Elgin, Cleavage of chromatin with methidiumpropyl-EDTA . iron(II). *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3213 (1983).
16. H. Ishii, J. T. Kadonaga, B. Ren, MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3457 (2015).
17. G. P. Oliveira *et al.*, The Investigation of Promoter Sequences of Marseilleviruses Highlights a Remarkable Abundance of the AAATATTT Motif in Intergenic Regions. *J. Virol.* **91**, (2017).
18. I. Pagnier *et al.*, A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. *Intervirology* **56**, 354 (2013).
19. R. V. Chereji, T. D. Bryson, S. Henikoff, Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol* **20**, 198 (2019).
20. S. K. Bhardwaj *et al.*, Dinucleosome specificity and allosteric switch of the ISW1a ATP-dependent chromatin remodeler in transcription regulation. *Nat Commun* **11**, 5913 (2020).

Acknowledgments: We thank Jim Kadonaga for MPE, Fred Hutch Genomics Shared Resource for sequencing services and Kami Ahmad for comments on the manuscript.

Funding: TDB, PBT and SH are supported by the Howard Hughes Medical Institute. PDI, MIV-S and K-JA were supported by a grant from the David and Lucile Packard Foundation.

Author contributions: Conceptualization: K-JA and SH; Investigation: TDB, PDI, MIV-S, JGH, PBT, BLS, K-JA, SH; Writing – original draft: SH; Writing – review & editing: TDB, PDI, MIV-S, JGH, PBT, K-JA, SH.

Competing interests: Authors declare that they have no competing interests.

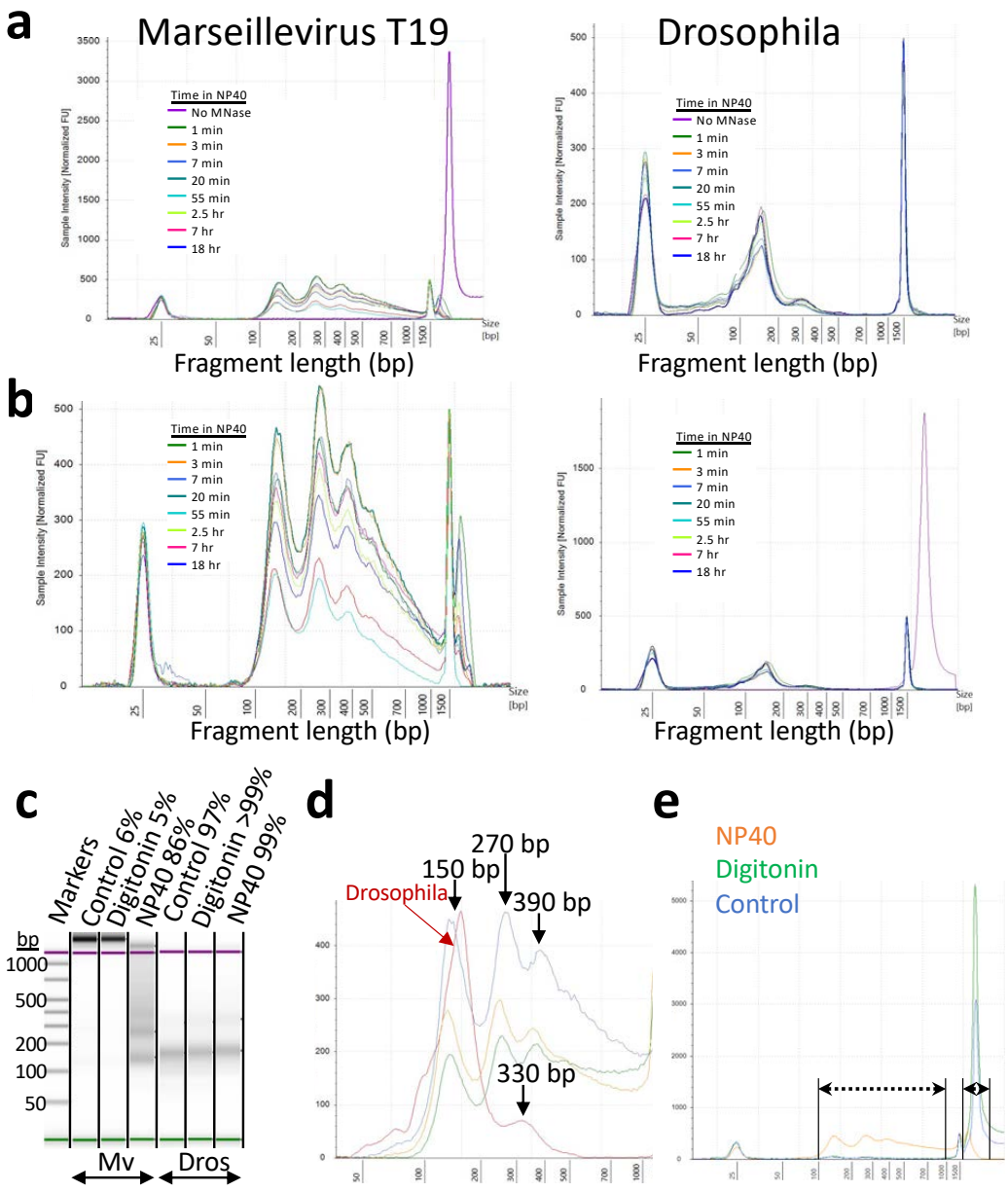

Figure S1: Permeabilization with NP-40 non-ionic detergent after capsid opening renders Marseillevirus chromatin immediately accessible to MNase digestion. a) Profile of Marseillevirus and Drosophila fragment lengths showing that NP40 completely permeabilizes pH2-treated and neutralized Marseillevirus particles to allow MNase digestion within 1 min. Estimated fragment sizes of peak midpoints are shown. **b)** Same as (a) showing that chromatin recovery is maximal in 1-3 minutes in NP40. **c)** Gel image of Marseillevirus (Mv) and Drosophila (Dros) chromatin extracted with no detergent (Control) or either Digitonin or NP-40 and digested with MNase at 10 U/million cells for 5 min at 37°C. **d)** Profiles of Mv and Dros fragment lengths showing recovery at 3 different concentrations. Estimated fragment sizes of peak midpoints are shown. **e)** Profiles of Mv lengths showing the ranges of signal used for calculating recovery: 100-1000 bp for MNase-released chromatin and >2 kb for undigested DNA.

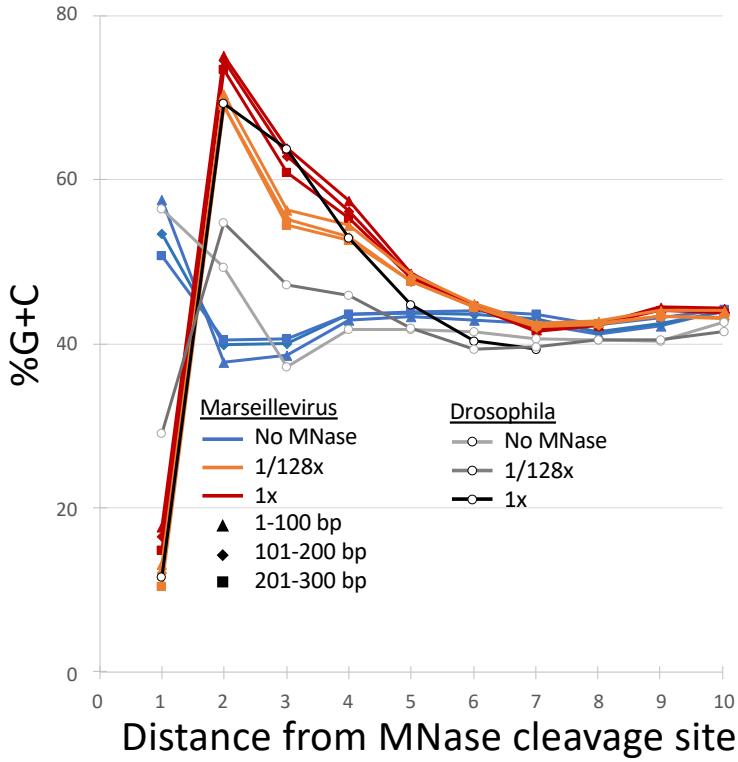

Figure S2: Preferential cleavage between A/T base pairs in Marseillevirus and Drosophila. MNase is an endo-exonuclease that cleaves preferentially at A/T-rich DNA then 'nibbles' on ends until it reaches a G/C-rich 'clamp', as observed in our MNase-seq data for both Marseillevirus and Drosophila, whose genomes are both 44-45% G+C overall.

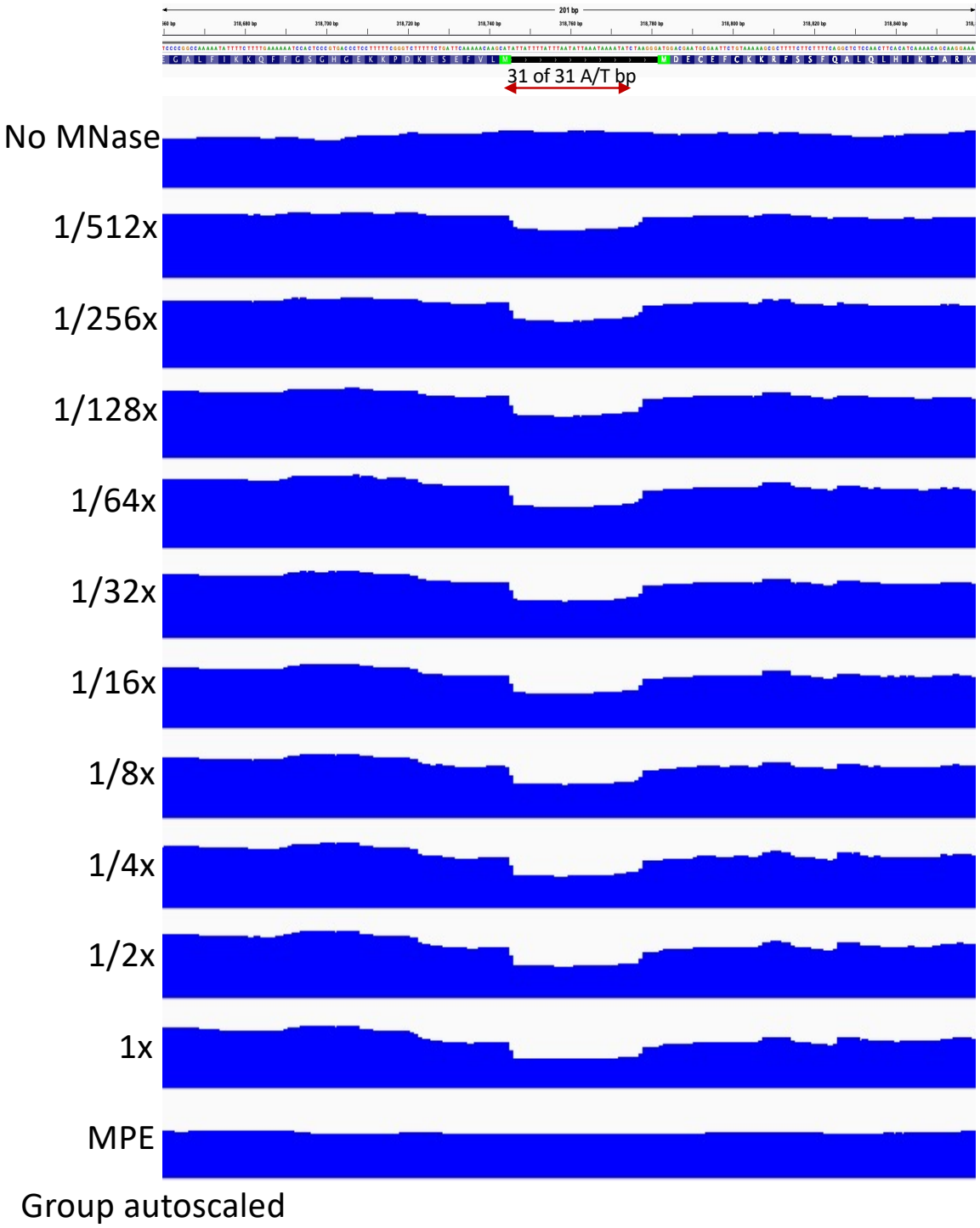

Figure S3: A/T-rich regions are preferentially digested by MNase. A Marseillevirus T19 bidirectional promoter with a 31/31 A/T bp span is preferentially digested by MNase over a 500-fold doubling series (1/512x,…,1x), whereas MPE shows no preference.

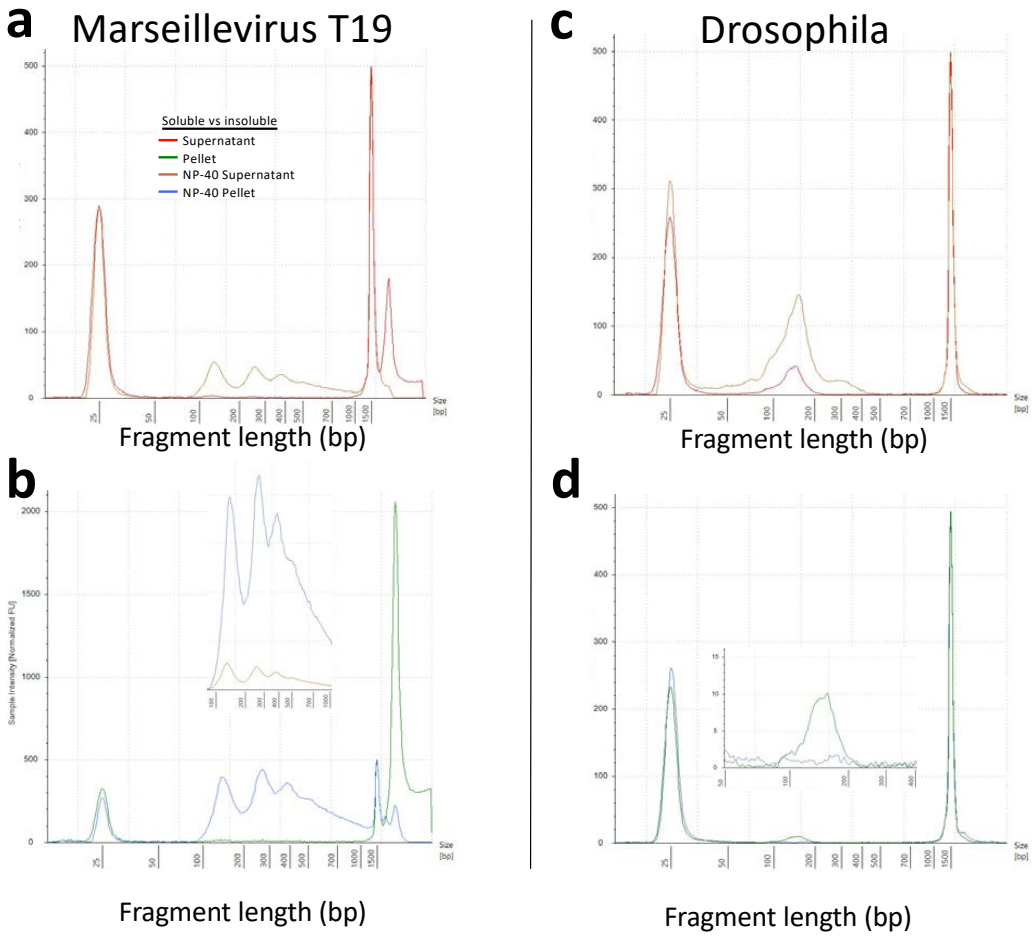

Figure S4: Marseillevirus nucleosomes are mostly insoluble after MNase treatment. Profiles of Marseillevirus T19 fragment lengths comparing ± NP-40 in centrifugation to separate **a)** supernatant and **b)** pellet, where the inset shows that most of the NP-40-released and digested Marseillevirus chromatin is in the insoluble pellet fraction with similar nucleosome profiles in the two fractions. **c-d)** In contrast, most Drosophila chromatin is soluble whether or not treated with NP-40, where the inset shows a magnification of the pellet profile, which reveals quantitative release of eukaryotic chromatin into the supernatant with NP-40.

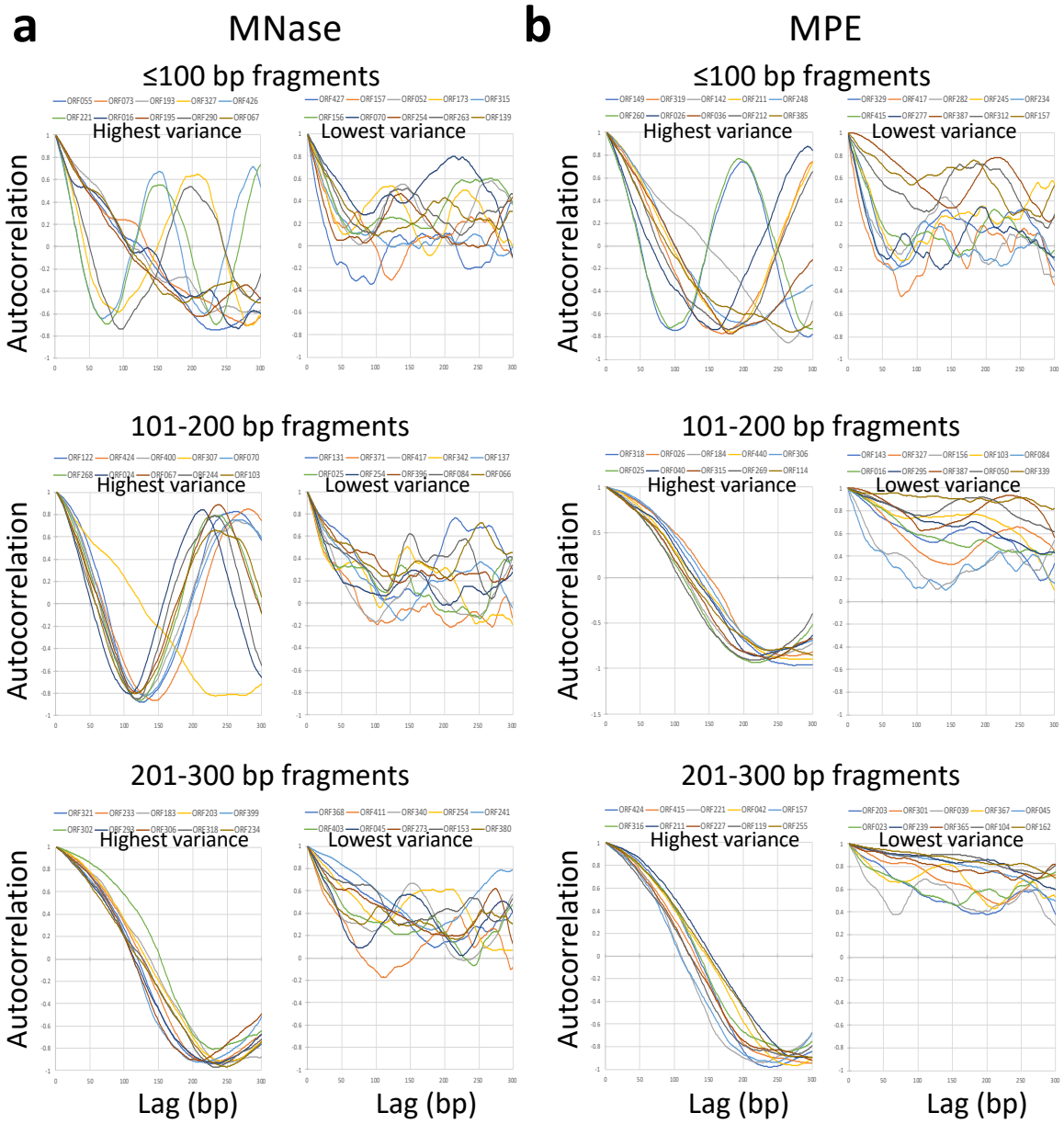

Figure S5: Autocorrelation illustrates periodicities over individual Marseillevirus gene bodies. To sensitively detect genes most likely to be phased, we plotted autocorrelations over the 600-bp span of each ORF for the 101-200 bp fragment size class using MNase-seq and MPE-seq data. Autocorrelations (-1 to +1) in 1-bp lag steps over a 5'-aligned 300-bp span are plotted for each of the 10 ORFs ≥600 bp with the highest and lowest variance in amplitude for each size class for **a**) MNase-seq and **b**) MPE-seq.