Supplementary Information for: Categorical Spectral Analysis of Periodicity in Nucleosomal DNA

Hu Jin, H. Tomas Rube, Jun S. Song

Contents

1 Supplementary Methods

1.1 Spectral envelope

1.1.1 Kernel smoothing of the periodogram

The calculation of sample spectral envelope (main text, *Materials and Methods*) involves estimating the power spectral density using periodogram. It has been suggested that smoothing the periodogram using a modified Daniell kernel provides a more stable estimate of the power spectral density [1]. We used a modified Daniell kernel $(0.125, 0.25, 0.25, 0.25, 0.125)$ for smoothing the periodogram in the calculation of sample spectral envelope. Specifically, the R code included in [2] smoothed the periodogram in the following way. For a numerical sequence of length L, the code calculated its periodogram $I(j/L)$ at each fundamental frequency j/L, where $j = -\lfloor L/2 \rfloor, -\lfloor L/2 \rfloor + 1, \ldots, -1, 0, 1, \ldots, \lfloor L/2 \rfloor - 1, \lfloor L/2 \rfloor$. The code then replaced $I(0)$ by $\frac{1}{2}(I(1/L) + I(-1/L))$ since $I(0)$ could be inflated by the mean of the sequence. It smoothed the periodogram by calculating $\hat{I}(j/L) = 0.125I(\frac{j-2}{L}) + 0.25I(\frac{j-1}{L}) + 0.25I(\frac{j}{L}) + 0.25I(\frac{j+1}{L}) + 0.125I(\frac{j+2}{L})$, where the arguments of I lying outside its domain of definition correspond cyclically to values in the interval $[-\lfloor L/2 \rfloor, \lfloor L/2 \rfloor]$. Because of the symmetry of periodogram, only $I(j/L)$ with $j > 0$ were retained in subsequent calculations. Notice that the smoothing introduced artifacts at the boundaries of the frequency space, specifically, at frequencies $1/L, 2/L, (\lfloor L/2 \rfloor - 1)/L,$ and $\lfloor L/2 \rfloor / L$. Thus the resulting sample spectral envelope might be inflated at these four frequencies. We therefore removed these four frequencies (or the corresponding periods) in all of the figures in frequency (period) space.

1.1.2 Empirical *p*-value

To assess the statistical significance of an observed periodic component from spectral envelope calculation, we defined an empirical p-value in the following way. Denote the spectral envelope at angular frequency ω of a DNA sequence s_t as $\lambda(\omega)$. We randomly permuted the sequence s_t 100 times and calculated the corresponding spectral envelope for each of the 100 permuted sequences, denoted as $\tilde{\lambda}_i(\omega)$, $i = 1, 2, \ldots, 100$. We then defined the empirical p-value as

$$p(\omega) = \frac{\sum_{i=1}^{100} \mathbb{1}_{\tilde{\lambda}_i(\omega) > \lambda(\omega)} + 1}{100 + 1}, \tag{S1}$$

where $\mathbb{1}$ is the indicator function. A pseudo-count 1 was added in both the numerator and denominator to avoid 0 p-values [3]. A small p-value indicates a significant periodic component.

To estimate the fraction of sequences that contained 10.5-bp periodicity, we fitted the distribution of empirical p-values (for example, Fig. 2B, solid black curve) using the following beta-uniform mixture model:

$$\Pr(p) = (1 - \pi_1) + \pi_1 \text{Beta}(p; 1, \beta), \tag{S2}$$

where the first component is the null uniform distribution between 0 and 1 with mixing coefficient $1 - \pi_1$, and the second component is a beta distribution with mixing coefficient π_1. $\text{Beta}(x; \alpha, \beta)$ denotes the probability density function of beta distribution with shape parameters α and β. We used the Expectation-Maximization (EM) algorithm to estimate the mixing coefficient π_1 and the shape parameter β. The estimated $\hat{\pi}_1$ characterizes the fraction of sequences containing 10.5-bp periodicity.

To correct for multiple hypothesis testing, we first calculated the 99% Clopper-Pearson confidence interval of each empirical p-value [4]. We then used the Monte Carlo Benjamini-Hochberg procedure to obtain an upper bound on the fraction of sequences that may contain statistically significant periodicity at 5% False Discovery Rate (FDR) [5, 6]. Specifically, the Clopper-Pearson confidence interval was defined as follows: for a specific DNA sequence, denote the total number of permutations with pseudo-count 1 as n (in our case, $n = 101$), and the total number of permuted sequences which have greater spectral envelope as k (in our case, $k = \sum_{i=1}^{n} \mathbb{1}_{\tilde{\lambda}_i(\omega) > \lambda(\omega)} + 1$). The lower and upper limits $[\hat{p}_l, \hat{p}_u]$ of the two sided $(1 - \alpha) \cdot 100\%$ Clopper-Pearson confidence interval of the empirical

p-value were defined [7] as

$$[\hat{p}_l, \hat{p}_u] = \begin{cases} \left[B(\frac{\alpha}{2}; k, n-k+1), B(1-\frac{\alpha}{2}; k+1, n-k)\right], & 1 \leq k < n \\ \left[\alpha^{1/(n+1)}, 1\right], & k = n \end{cases} \tag{S3}$$

where $B(q; a, b)$ is the qth quantile of the beta distribution with shape parameters a and b. We used $\alpha = 0.01$ to obtain 99% confidence intervals. The traditional Benjamini-Hochberg procedure for controlling FDR works as follows: given N p-values p_i, $i = 1, 2, \ldots, N$, denote their order statistic as $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(N)}$. Suppose j is the largest index i for which $p_{(i)} \leq \frac{i}{N}\theta$; then, rejecting all the hypotheses corresponding to $p_{(1)}, p_{(2)}, \ldots, p_{(j)}$ guarantees that the FDR is at most θ. To obtain an upper bound on the fraction of rejected hypotheses at fixed FDR θ, we compare the lower limits $\hat{p}_l$ of the Clopper-Pearson confidence interval with the Benjamini-Hochberg line $\frac{i}{N}\theta$. Denote the lower limits of the Clopper-Pearson confidence intervals corresponding to the empirical p-values p_i as $\hat{p}_{l,i}$, and their order statistic as $\hat{p}_{l,(i)}$. Suppose j is the smallest index i for which the condition $\hat{p}_{l,(k)} > \frac{k}{N}\theta$ holds for all $\hat{p}_{l,(k)} \geq \hat{p}_{l,(i)}$. Then, at most $j-1$ hypotheses corresponding to $\hat{p}_{l,(1)}, \hat{p}_{l,(2)}, \ldots, \hat{p}_{l,(j-1)}$ may be rejected at this chosen FDR, and $1 - j/N$ provides an upper bound on the fraction of rejected hypotheses at FDR θ. We used FDR $\theta = 5\%$.

1.1.3 Computation details

We used a modified version of the R code included in [2] to compute the sample spectral envelope. The R code was called from a Python wrapper using rpy2 (`http://rpy.sourceforge.net/`). Parallelization was implemented using PP (`http://www.parallelpython.com/`). Clopper-Pearson confidence intervals were calculated in R using the package GenBinomApps (`https://cran.r-project.org/web/packages/GenBinomApps/index.html`).

1.2 Spurious periodicity

We simulated 10,000 147-bp random sequences with the expected frequency of A, C, G, and T set to 0.3, 0.2, 0.2, and 0.3, respectively. We then converted these sequences into numerical sequences by setting A and T to 1, C and G to 0, and took the average of these 10,000 numerical sequences at each of the 147 positions. Equivalently, we were counting the average A/T frequency at each position of the center-aligned sequences (grey curve in Fig. 1B).

To create a spurious 10-bp periodicity in the average A/T frequency, we performed the following re-alignment of the sequences [8]. For each of the 10,000 simulated sequences, we computed the mode of the discrete probability distribution of the positions of A/T modulo 10, denoted as m_i, where $i = 1, \ldots, 10,000$ is the index of sequences, and $m_i \in \{0, 1, \ldots, 9\}$. We then cyclically shifted the ith sequence by $m_i + \delta_i$, where δ_i is a small noise introduced to simulate imperfect alignment among nucleosomal sequences. We simulated δ_i as a discrete random variable taking values $-2, -1, 0, 1, 2$ with probability $0.125, 0.25, 0.25, 0.25, 0.125$, respectively. Finally, we computed the average A/T frequency at each position in the alignment of shifted sequences (black curve in Fig. 1B).

A modification of this method can produce arbitrary periodicity in random sequences. These results demonstrate that periodicity in average nucleotide frequency is not equivalent to periodicity in individual sequences.

1.3 Spectral decomposition

In this section, we describe in detail the spectral decomposition method for quantifying the origin of periodicity observed in average nucleotide frequency of aligned nucleosomal sequences.

1.3.1 The decomposition equation

Consider N nucleosomal sequences $s_k(t)$ of length L (typically $L = 147$), for $k = 1, 2, \ldots, N$ and $t = 0, 1, \ldots, L-1$. The average nucleotide frequency (for example, as shown in Fig. 1A) can be calculated as follows. We may choose a specific scaling function β and map the nucleosomal sequences to real-valued sequences, and then take the average across the N numerical sequences. For example, if we are interested in the average A/T frequency (as in Fig. 1A), we may choose the scaling function to be $\beta : (\text{A, C, G, T}) \to (1, 0, 0, 1)$. The real-valued sequences obtained using the scaling function is denoted as $x_k(t) = \beta(s_k(t))$. The average nucleotide frequency at location t is simply the average of $x_k(t)$ across nucleosomes; i.e., $\bar{x}(t) = \frac{1}{N}\sum_{k=1}^{N} x_k(t)$. A dominant periodic component in average nucleotide frequency can be indicated by a peak in the spectrum of spectral density of $\bar{x}(t)$, which is the squared norm of the discrete Fourier transform of $\bar{x}(t)$

$$|\tilde{\bar{x}}(f)|^2 = \left|\frac{1}{\sqrt{L}}\sum_{t=0}^{L-1}\bar{x}(t)e^{-2\pi i f t}\right|^2, \tag{S4}$$

where $\tilde{x}$ denotes the discrete Fourier transform of x, and f is the fundamental frequency, where $f = j/L$, $j = 0, 1, 2, \ldots, (L-1)/2$ for odd L, and $j = 0, 1, 2, \ldots, L/2$ for even L. Therefore, $|\tilde{\bar{x}}(f)|^2$ assesses the periodicity in average nucleotide frequency of aligned nucleosomal sequences. We will decompose $|\tilde{\bar{x}}(f)|^2$ into several factors that are easy to interpret.

Since Fourier transform is a linear operator, the order of taking average and taking discrete Fourier transform in Equation S4 is exchangeable, and we can rewrite

$$|\tilde{\bar{x}}(f)|^2 = \left|\frac{1}{N}\sum_{k=1}^{N}\tilde{x}_k(f)\right|^2 = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}|\tilde{x}_k(f)|\,|\tilde{x}_l(f)|\,e^{i(\phi_k(f)-\phi_l(f))}, \tag{S5}$$

where $\phi_k(f) = \text{Arg}(\tilde{x}_k(f))$ is the argument of the discrete Fourier transform. By simply multiplying terms on both sides of Equation S5, we obtained the following decomposition equation

$$\frac{|\tilde{\bar{x}}|^2}{\frac{1}{N}\frac{1}{N}\sum_m \text{E}\left[|\tilde{x}_m|^2\right]} = \frac{\frac{1}{N}\sum_m|\tilde{x}_m|^2}{\frac{1}{N}\sum_m \text{E}\left[|\tilde{x}_m|^2\right]}\,\frac{\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}}{\frac{1}{N^2}\sum_k\sum_l \text{E}\left[e^{i(\phi_k-\phi_l)}\right]}\,\frac{\frac{1}{N^2}\sum_k\sum_l \frac{|\tilde{x}_k||\tilde{x}_l|}{\frac{1}{N}\sum_m|\tilde{x}_m|^2}e^{i(\phi_k-\phi_l)}}{\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}}, \tag{S6}$$

where the argument $f \neq 0$ was omitted for all variables for clarity, the summations over k, l, m are from 1 to N, E denotes expectation over all possible independent permutations of the N nucleosomal sequences $s_k(t)$, i.e. averaging over permuting $s_k(t)$ with independent uniformly sampled elements π_k of the symmetric group S_L, and the following theorem has been used.

Theorem 1.1 *Under the settings developed above, at frequency $f \neq 0$,*

$$\frac{1}{N^2}\sum_k\sum_l E\left[e^{i(\phi_k-\phi_l)}\right] = \frac{1}{N}. \tag{S7}$$

Proof If $k = l$, then $e^{i(\phi_k-\phi_l)} = e^0 = 1$. Thus, $\text{E}\left[e^{i(\phi_k-\phi_l)}\right] = 1$ if $k = l$. If $k \neq l$, since permutations are carried out independently on sequences s_k and s_l, we have $\text{E}\left[e^{i(\phi_k-\phi_l)}\right] = \text{E}\left[e^{i\phi_k}\right]\text{E}\left[e^{-i\phi_l}\right]$. If for any k, $\text{E}\left[e^{i\phi_k}\right] = \text{E}\left[e^{-i\phi_k}\right] = 0$, then $\frac{1}{N^2}\sum_k\sum_l \text{E}\left[e^{i(\phi_k-\phi_l)}\right] = \frac{1}{N^2}\sum_k \text{E}\left[e^{i(\phi_k-\phi_k)}\right] + \frac{1}{N^2}\sum_k\sum_{l\neq k}\text{E}\left[e^{i(\phi_k-\phi_l)}\right] = \frac{1}{N^2}N + 0 = \frac{1}{N}$, and the theorem follows. The following lemma guarantees that for any k, $\text{E}\left[e^{i\phi_k}\right] = \text{E}\left[e^{-i\phi_k}\right] = 0$.

Lemma 1.2 *Under the settings developed above, at frequency $f \neq 0$, for any k, $E\left[e^{i\phi_k}\right] = E\left[e^{-i\phi_k}\right] = 0$.*

Proof Since $E\left[e^{-i\phi_k}\right]$ is complex conjugate of $E\left[e^{i\phi_k}\right]$, we only need to show that $E\left[e^{i\phi_k}\right] = 0$. Let $\mathbb{Z}_L \subset S_L$ be the cyclic subgroup that cyclically permutes a sequence. Then, we can partition S_L into equivalence classes defined by the right cosets $\mathbb{Z}_L\pi$, $\pi \in S_L$. We will show that the sum of the phases of the permuted sequences obtained from each coset vanishes. For any sequence s, the action of the coset $\mathbb{Z}_L\pi$ on s is a constant translation of the permuted sequence $s' = \pi(s)$, where the translation is defined on a periodic lattice of length L. Let $g \in \mathbb{Z}_L$ be the generator that translates the sequence by one unit; note that $\mathbb{Z}_L = \{g, g^2, \ldots, g^L\}$. Let x' denote the numerical representation of s', and $\tilde{x}'$ its Fourier transform. Since the Fourier dual of translation is a phase shift, the Fourier transform of the translated sequence $g^j s'$ is $\widetilde{(g^j x')} = \omega^j \tilde{x}'$, where $\omega = e^{-2\pi i/L}$ is an L-th root of unity. Thus, for any $\pi \in S_L$,

$$\sum_{h\in\mathbb{Z}_L\pi} \widetilde{hx} = \sum_{j=0}^{L-1} \widetilde{g^j x'} = \tilde{x}' \sum_{j=0}^{L-1} \omega^j = \tilde{x}' \frac{1-\omega^L}{1-\omega} = 0,$$

for $\omega \neq 1$, i.e. for $f \neq 0$. Since for any fixed $\pi \in S_L$, $\widetilde{hx}$ have the same magnitude for all $h \in \mathbb{Z}_L\pi$, the claim thus follows. ▮

1.3.2 Interpretation of each factor

We proceed to define *aligned enrichment* A as the left hand side of the decomposition Equation S6,

$$A = \frac{|\tilde{\tilde{x}}|^2}{\frac{1}{N}\frac{1}{N}\sum_m \mathrm{E}\left[|\tilde{x}_m|^2\right]}, \tag{S8}$$

individual enrichment I as the first factor on the right hand side of Equation S6,

$$I = \frac{\frac{1}{N}\sum_m |\tilde{x}_m|^2}{\frac{1}{N}\sum_m \mathrm{E}\left[|\tilde{x}_m|^2\right]}, \tag{S9}$$

phasing enrichment P as the second factor on the right hand side of Equation S6,

$$P = \frac{\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}}{\frac{1}{N^2}\sum_k\sum_l \mathrm{E}\left[e^{i(\phi_k-\phi_l)}\right]}, \tag{S10}$$

and the *residual factor* R as the third factor on the right hand side of Equation S6,

$$R = \frac{\frac{1}{N^2}\sum_k\sum_l \frac{|\tilde{x}_k||\tilde{x}_l|}{\frac{1}{N}\sum_m|\tilde{x}_m|^2} e^{i(\phi_k-\phi_l)}}{\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}}. \tag{S11}$$

The decomposition Equation S6 can be then simplified as

$$A = IPR, \tag{S12}$$

which is reminiscent of Equation 2 in the main text.

Equation S9 can be further expressed as

$$I = \frac{\frac{1}{N}\sum_m |\tilde{x}_m|^2}{\mathrm{E}\left[\frac{1}{N}\sum_m |\tilde{x}_m|^2\right]}, \tag{S13}$$

since $\frac{1}{N}\sum_m \mathrm{E}\left[|\tilde{x}_m|^2\right] = \mathrm{E}\left[\frac{1}{N}\sum_m |\tilde{x}_m|^2\right]$. Similarly, Equation S10 can be further expressed as

$$P = \frac{\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}}{\mathrm{E}\left[\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}\right]}, \tag{S14}$$

because $\frac{1}{N^2}\sum_k\sum_l \mathrm{E}\left[e^{i(\phi_k-\phi_l)}\right] = \mathrm{E}\left[\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)}\right]$. Equation S8 can be further expressed as

$$A = \frac{|\tilde{\tilde{x}}|^2}{\mathrm{E}\left[|\tilde{\tilde{x}}|^2\right]}, \tag{S15}$$

because of the following lemma.

Lemma 1.3 *Under the settings developed above,* $\frac{1}{N}\frac{1}{N}\sum_m E\left[|\tilde{x}_m|^2\right] = E\left[|\tilde{\bar{x}}|^2\right]$*, for frequency* $f \neq 0$.

Proof Taking expectation of both sides of Equation S5, we have

$$\begin{aligned}
\mathrm{E}\left[|\tilde{\bar{x}}|^2\right] &= \frac{1}{N^2}\mathrm{E}\left[\sum_{k=1}^{N}\sum_{l=1}^{N}|\tilde{x}_k|\,|\tilde{x}_l|\,e^{i(\phi_k-\phi_l)}\right] \\
&= \frac{1}{N^2}\sum_{k=l}\mathrm{E}\left[|\tilde{x}_k|\,|\tilde{x}_l|\,e^{i(\phi_k-\phi_l)}\right] + \frac{1}{N^2}\sum_{k\neq l}\mathrm{E}\left[|\tilde{x}_k|\,|\tilde{x}_l|\,e^{i(\phi_k-\phi_l)}\right] \\
&= \frac{1}{N^2}\sum_{m}\mathrm{E}\left[|\tilde{x}_m|^2\right] + \frac{1}{N^2}\sum_{k\neq l}\mathrm{E}\left[|\tilde{x}_k|\,|\tilde{x}_l|\,e^{i(\phi_k-\phi_l)}\right].
\end{aligned} \tag{S16}$$

We shall show that $\frac{1}{N^2}\sum_{k\neq l}\mathrm{E}\left[|\tilde{x}_k|\,|\tilde{x}_l|\,e^{i(\phi_k-\phi_l)}\right] = 0$, from which the lemma follows. Due to the independence of permutations for different sequences, we have

$$\frac{1}{N^2}\sum_{k\neq l}\mathrm{E}\left[|\tilde{x}_k|\,|\tilde{x}_l|\,e^{i(\phi_k-\phi_l)}\right] = \frac{1}{N^2}\sum_{k\neq l}\mathrm{E}\left[\tilde{x}_k\right]\mathrm{E}\left[\tilde{x}_l^*\right], \tag{S17}$$

where $*$ denotes complex conjugation. It thus suffices to show that for any k, $\mathrm{E}\left[\tilde{x}_k\right] = \mathrm{E}\left[\tilde{x}_k^*\right] = 0$, but they were already proved in Lemma 1.2. $\blacksquare$

The *aligned enrichment* A (Equation S15, Lemma 1.3) is the spectral density of average nucleotide frequency divided by its expectation. It thus quantifies the *spectral density enrichment of average nucleotide frequency* compared to randomly permuted sequences.

The *individual enrichment* I (Equation S13) is the average spectral density of individual nucleosomal sequences divided by its expectation. It thus quantifies the *spectral density enrichment of individual nucleosomal sequences* compared to randomly permuted sequences.

The *phasing enrichment* P (Equation S14) is also of the form that the denominator is the expectation of the numerator. Specifically, the numerator is 1 when all of the sequences are in phase with each other, i.e., $\phi_k = \phi_l$ for any k and l. When the sequences are not completely in phase with each other, the numerator is real and always smaller than 1, because

$$\begin{aligned}
\frac{1}{N^2}\sum_k\sum_l e^{i(\phi_k-\phi_l)} &= \frac{1}{N^2}\sum_k\sum_l \mathrm{Re}(e^{i(\phi_k-\phi_l)}) \\
&= \frac{1}{N^2}\sum_k\sum_l \cos(\phi_k-\phi_l) \\
&\leq \frac{1}{N^2}\sum_k\sum_l 1 \\
&= 1,
\end{aligned} \tag{S18}$$

where the equality holds if and only if $\phi_k = \phi_l, \forall k \neq l$. Therefore, the numerator in Equation S14 is a measure of the degree of phasing between nucleosomal sequences. The *phasing enrichment* P thus quantifies the *enrichment of phasing among nucleosomal sequences* compared to randomly permuted sequences.

The *residual factor* R (Equation S11) is subjected to random noise and does not contain useful information about the decomposition, as discussed in the main text.

1.3.3 Calculation details

Each factor in the decomposition Equation S6 can be calculated given a fixed scaling function β, without having to generate all permutations of a nucleosomal sequence, if the term $\mathrm{E}\left[|\tilde{x}_m|^2\right]$ can be explicitly calculated. We calculate this term for the specific scaling function $\beta : (\mathrm{A, C, G, T}) \to (1,0,0,1)$ used in the main text. The calculation can be generalized easily to cases where other scaling functions are used. We have

$$\begin{aligned}
\mathrm{E}\left[|\tilde{x}_m|^2\right] &= \mathrm{E}\left[\frac{1}{L}\sum_{t=0}^{L-1}\sum_{t'=0}^{L-1} x_m(t)x_m(t')e^{-2\pi i f(t-t')}\right] \\
&= \frac{1}{L}\sum_{t=0}^{L-1}\mathrm{E}\left[x_m(t)^2\right] + \frac{1}{L}\sum_{t\neq t'}\mathrm{E}\left[x_m(t)x_m(t')\right]e^{-2\pi i f(t-t')}.
\end{aligned} \tag{S19}$$

Under permutation, each position $t = 0, 1, \ldots, L-1$ is equivalent. Thus for any t, $\mathrm{E}\left[x_m(t)^2\right] = \mathrm{E}\left[x_m(0)^2\right]$. Moreover, since each mono-nucleotide is independently mapped to a real number according to the scaling function $\beta : (\mathrm{A, C, G, T}) \to (1,0,0,1)$, we have for any $t \neq t'$, $\mathrm{E}\left[x_m(t)x_m(t')\right] = \mathrm{E}\left[x_m(0)x_m(1)\right]$. Denote the number of A and T's in the nucleosomal sequence s_m as r_m, i.e., r_m is the number of 1's in the numerical sequence x_m. When $r_m \geq 1$, we have $\mathrm{E}\left[x_m(0)^2\right] = \frac{r_m}{L}$ and $\mathrm{E}\left[x_m(0)x_m(1)\right] = \frac{r_m}{L}\frac{r_m-1}{L-1}$. When $r_m = 0$, we have $\mathrm{E}\left[x_m(0)^2\right] = 0$ and $\mathrm{E}\left[x_m(0)x_m(1)\right] = 0$. Furthermore, for $f \neq 0$,

$$\begin{aligned}
\frac{1}{L}\sum_{t\neq t'} e^{-2\pi i f(t-t')} &= \frac{1}{L}\sum_{t,t'} e^{-2\pi i f(t-t')} - \frac{1}{L}\sum_{t=t'} e^{-2\pi i f(t-t')} \\
&= \frac{1}{L}\left|\sum_{t=0}^{L-1} e^{-2\pi i f t}\right|^2 - 1 \\
&= -1.
\end{aligned} \tag{S20}$$

Therefore, for $f \neq 0$, we have shown

$$\mathrm{E}\left[|\tilde{x}_m|^2\right] = \frac{r_m(L-r_m)}{L(L-1)},$$

where r_m is the number of A and T's in the nucleosomal sequence s_m.

1.4 Circular-linear and circular-circular correlation

Given i.i.d. bivariate data (α_k, x_k), $k = 1, 2, \ldots, N$, where α_k is circular data in $[0, 2\pi)$, and x_k is linear data in $\mathbb{R}$, the circular-linear correlation between α and x is defined [9] as

$$r^2 = \frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2}, \tag{S21}$$

where $r_{xs} = \mathrm{corr}\,(x, \sin\alpha)$, $r_{xc} = \mathrm{corr}\,(x, \cos\alpha)$, $r_{cs} = \mathrm{corr}\,(\cos\alpha, \sin\alpha)$, and corr is the sample Pearson correlation coefficient. This circular-linear correlation was used in Fig. 5B and Fig. S7B to characterize the correlation between the strength of 10.5-bp periodicity in nucleosomal sequences and the degree of phasing among these sequences. For this purpose, we first mapped each nucleosomal sequence s_k to a real-valued sequence x_k using a fixed scaling function, which is $\beta : (\mathrm{A, C, G, T}) \to (1,0,0,1)$ in the main text. We then calculated the discrete Fourier transform $\tilde{x}_k$. The circular-linear correlation was calculated between the circular data $\phi_k = \mathrm{Arg}(\tilde{x}_k)$ and linear data $\sqrt{\frac{2|\tilde{x}_k|^2}{\mathrm{var}(x_k)}}$ at some fixed frequency f, where var is the sample variance and $\frac{2|\tilde{x}_k|^2}{\mathrm{var}(x_k)}$ is the power spectral density normalized by total power. In Fig. 5A and Fig. S7A, we ranked the nucleosomes according to the strength of 10.5-bp periodicity of their underlying sequences (the power spectral density normalized by total power $\frac{2|\tilde{x}_k|^2}{\mathrm{var}(x_k)}$) and divided them into 5 quintiles of equal size (the five circular rings in Fig. 5A and Fig. S7A). We then binned the phases of their discrete Fourier transform at period 10.5 bp ($\phi_k(f = 1/10.5)$) into 20 equal intervals in $[0, 2\pi)$ (the 20 bins within each ring in Fig. 5A and Fig. S7A). The fraction of nucleosomes lying in each bin within the corresponding quintile divided by the expectation under uniform distribution is indicated by the color.

Notice that in Fig. 5A, Fig. S7A and Fig. S8, we chose to count the phase with respect to the dyad. Specifically, the phase of nucleosomal sequence s_k with respect to the dyad at frequency f was defined as $\text{Arg}(\tilde{x}_k(f)e^{2\pi if\times 73})$, where the extra phase factor $e^{2\pi if\times 73}$ was introduced to enforce the origin to be at the dyad. Equivalently, one may calculate the discrete Fourier transform using a translated nucleotide index $\frac{1}{\sqrt{L}}\sum_{t=-73}^{73} x_k(t)e^{-2\pi ift}$ and then take its argument. Note that the circular-linear correlation is invariant with respect to the choice of the origin of phase.

To determine whether the phase of A/T nucleotides in nucleosomal sequences agrees with the phase of histone-DNA contact points, we created an indicator vector of length 147, where we put 1 at histone-DNA contact points and 0 at other locations. The locations of histone-DNA contact points were taken to be the sites of primary bound-phosphate groups that show conserved interaction with histones, as determined by the crystal structure of nucleosome [10]; the locations are at $-55, -44, -34, -24, -13, -3, 3, 13, 24, 34, 44, 55$ bp with respect to the dyad axis (chosen to be 0). We then performed discrete Fourier transform of this indicator vector and determined its phase at period 10.5 bp with respect to the dyad axis, as described in the previous paragraph. The phase of histone-DNA contact points at period 10.5 bp with respect to the dyad axis was calculated to be 3.12.

Given i.i.d. bivariate data (α_k, β_k), $k = 1, 2, \ldots, N$, where both α_k and β_k are circular data in $[0, 2\pi)$, the circular-circular correlation coefficient between α and β is defined [9] as

$$r = \frac{\sum_{k=1}^{N} \sin(\alpha_k - \bar{\alpha})\sin(\beta_k - \bar{\beta})}{\sqrt{\sum_{k=1}^{N} \sin^2(\alpha_k - \bar{\alpha})\sin^2(\beta_k - \bar{\beta})}}, \tag{S22}$$

where $\bar{\alpha}$ and $\bar{\beta}$ are the mean directions of α_k and β_k, respectively. The mean direction $\bar{\alpha}$ of circular data α_k is defined [9] as

$$\bar{\alpha} = \arctan^*(S/C) = \begin{cases} \arctan(S/C), & \text{if } C > 0, S \geq 0, \\ \pi/2, & \text{if } C = 0, S > 0, \\ \arctan(S/C) + \pi & \text{if } C < 0, \\ \arctan(S/C) + 2\pi & \text{if } C \geq 0, S < 0, \\ \text{undefined} & \text{if } C = 0, S = 0, \end{cases} \tag{S23}$$

where $S = \sum_{k=1}^{N} \sin\alpha_k$ and $C = \sum_{k=1}^{N} \cos\alpha_k$. A similar definition holds for $\bar{\beta}$. This circular-circular correlation coefficient was used in Fig. 5C and Fig. S9 to characterize the phasing between nucleosome position and the 10.5-bp periodicity of the underlying sequence. To do that, we first mapped each consensus nucleosomal sequence s_k centered at genomic location u_k to a real-valued sequence x_k using a fixed scaling function, which is β : (A, C, G, T) $\rightarrow$ $(1, 0, 0, 1)$ in the main text. We then calculated the discrete Fourier transform $\tilde{x}_k$ and took the argument $\phi_k(f = 1/10.5) = \text{Arg}(\tilde{x}_k(f = 1/10.5))$ of the discrete Fourier transform at period 10.5 bp as one of the circular data sets (for example, α_k in Equation S22). For the other circular data set (for example, β_k in Equation S22), we created an indicator vector $y_k(t)$, $t = 0, 1, \ldots, 146$ of length 147 for each consensus nucleosome s_k, where $y_k(t) = 1$ if there is a redundant nucleosome centered at genomic location $u_k - 73 + t$, and $y_k(t) = 0$ otherwise. y_k is thus a vector indicating all possible positions that the consensus nucleosome may occupy in cell population. We then calculated the discrete Fourier transform $\tilde{y}_k$ and took its argument $\psi_k(f = 1/10.5) = \text{Arg}(\tilde{y}_k(f = 1/10.5))$ at period 10.5 bp as the other circular data set (for example, β_k in Equation S22). The circular-circular correlation coefficient was calculated between the bivariate data (ϕ_k, ψ_k). Specifically, we restricted ourselves to the consensus nucleosomes with at least 5 redundant nucleosomes lying within ±73 bp of its dyad position, i.e., $\sum_{t=0}^{146} y_k(t) \geq 5$ (34,020 out of 67,531 in *S. cerevisiae* and 41,957 out of 75818 in *S. pombe*. These numbers are different from those in section 1.5, because we used the window size ±60 bp in that section as suggested by [11], while here we used the window size ±73 bp to obtain an indicator sequence of the same length as a nucleosomal sequence). In Fig. 5C and Fig. S9, we first ranked the consensus nucleosomes (which satisfy $\sum_{t=0}^{146} y_k(t) \geq 5$) by the strength of 10.5-bp periodicity of their underlying sequences (the power spectral density normalized by total power $\frac{2|\tilde{x}_k|^2}{\text{var}(x_k)}$), from small to large values. We then divided the consensus nucleosomes into 10 groups of equal size and calculated the circular-circular correlation coefficient between ϕ_k and ψ_k within each group. The correlation coefficient was plotted against the group index, with group 1 corresponding to the smallest strength of 10.5-bp periodicity in nucleosomal sequences (black curves in Fig. 5C and Fig. S9). We also performed the same analysis for other fundamental frequencies. For each group index $1, 2, \ldots, 10$, we plotted the median correlation coefficient among all fundamental frequencies excluding $1/10.5$, with whiskers showing the range from the 5th percentile to the 95th percentile (blue curves in Fig. 5C and Fig. S9).

1.5 Fuzziness scores

Nucleosome positioning and nucleosome occupancy are two related but distinct concepts. Nucleosome positioning can be further distinguished as translational and rotational positioning. We thus clearly define these concepts before proceeding to define appropriate scores quantifying these different aspects.

Nucleosome occupancy refers to the probability that a given base pair in the genome is occupied by a nucleosome in cell population [12, 13], while nucleosome positioning quantifies the degree to which the position of an individual nucleosome varies across the cell population [13, 14]. Thus, nucleosome occupancy characterizes the level of nucleosome depletion at a genomic location, while nucleosome positioning measures how well positioned a nucleosome is.

For an individual nucleosome, translational positioning refers to the degree to which the location of the 147-bp DNA contacting histone octamer varies in cell population, while rotational positioning refers to the degree to which the rotational orientation of DNA helix relative to the histone surface varies in cell population [12]. In this section, we define two scores termed the translational and rotational fuzziness to quantify translational and rotational positioning, respectively. Previous studies utilizing the chemical cleavage method to map genome-wide nucleosome locations [15, 11] yielded redundant maps of nucleosomes, representing all possible nucleosome positions across cell population, along with unique maps representing consensus nucleosome positions. We will now define the translational and rotational fuzziness of a consensus nucleosome in the unique map by assessing the translational and rotational variance of redundant nucleosome positions associated with the consensus nucleosome.

1.5.1 Translational fuzziness

Denote the set of dyad positions of consensus nucleosomes from the unique map as $\mathcal{N}_U$, and the set of dyad positions of redundant nucleosomes as $\mathcal{N}_R$. For a consensus nucleosome $u \in \mathcal{N}_U$, suppose there are n_u redundant nucleosomes $r_1, r_2, \ldots, r_{n_u} \in \mathcal{N}_R$ that are within ±60 bp of u, i.e., $|r_i - u| \leq 60$, $i = 1, 2, \ldots, n_u$. We assume that these n_u positions are all possible positions that the consensus nucleosome u may occupy in different cells. Each position r_i has a nucleosome center positioning (NCP) score k_{r_i} [15, 16], measuring the relative number of nucleosomes centered at r_i in cell population. We should therefore think of all possible positions that the consensus nucleosome u may occupy as r_1 with probability proportional to k_{r_1}, r_2 with probability proportional to k_{r_2}, and so on. Denote the relative coordinates of r_i with respect to u as $x_{i,u} = r_i - u$. We assume that the expectation value of $x_{i,u}$ is 0, meaning that the expected position of the consensus nucleosome u is u. We define the translational fuzziness $F_u^{(trans.)}$ of the consensus nucleosome u as the variance of the n_u relative coordinates $x_{i,u}$, weighted by the NCP score,

$$F_u^{(trans.)} = \frac{\sum_{i=1}^{n_u} k_{r_i} x_{i,u}^2}{\sum_{i=1}^{n_u} k_{r_i}}. \tag{S24}$$

Therefore, small translational fuzziness indicates a high degree of translational positioning of a consensus nucleosome, and the nucleosome is said to be well positioned translationally in this case. Notice that when n_u is small, we are not able to obtain a good estimate of the variance. We thus calculated translational fuzziness for only those consensus nucleosomes with $n_u \geq 5$ (30,628 out of 67,531 nucleosomes in *S. cerevisiae* and 40,384 out of 75,818 nucleosomes in *S. pombe*).

1.5.2 Rotational fuzziness

Using the same notation as above, the rotational fuzziness $F_u^{(rot.)}$ of a consensus nucleosome u can be defined as follows. To capture the relative orientation of DNA helix relative to the histone surface, we converted the relative coordinates $x_{i,u}$ to relative circular coordinates $\theta_{i,u}^{(h)} = \frac{x_{i,u} \bmod h}{h} \times 2\pi$, where h is the helical repeat length. The rotational fuzziness of the consensus nucleosome u is then defined as the circular variance [9] of these relative circular coordinates, weighted by NCP score,

$$\hat{F}_u^{(rot.)}(h) = 1 - \frac{1}{\sum_{i=1}^{n_u} k_{r_i}} \sqrt{\left(\sum_{i=1}^{n_u} k_{r_i} \cos\theta_{i,u}^{(h)}\right)^2 + \left(\sum_{i=1}^{n_u} k_{r_i} \sin\theta_{i,u}^{(h)}\right)^2}. \tag{S25}$$

Since the helical repeat length h within a nucleosome may vary across different locations and may be different from that of a free DNA molecule [17], we finally defined the rotational fuzziness as

$$F_u^{(rot.)} = \min_{h\in\{9.5,9.6,\ldots,11.5\}} \hat{F}_u^{(rot.)}(h). \tag{S26}$$

Thus, small rotational fuzziness indicates a high degree of rotational positioning of a consensus nucleosome, and the nucleosome is said to be well positioned rotationally in this case. Because of the same reason mentioned in the definition of translational fuzziness, we defined rotational fuzziness for only those consensus nucleosomes with $n_u \geq 5$.

1.6 Nucleosome occupancy

We followed [11] for the definition of nucleosome occupancy. For both *S. cerevisiae* and *S. pombe*, denote the set of nucleosome dyad positions in the redundant map as $\mathcal{N}_R$. The nucleosome occupancy O_i of a genomic location i is defined as the total NCP score of the redundant nucleosomes in the ±60-bp region of i,

$$O_i = \sum_{|j-i|\leq 60,\ j\in\mathcal{N}_R} k_j, \tag{S27}$$

where k_j is the NCP score of genomic location j. The NCP score k_j can be interpreted as the relative number of nucleosomes centered at genomic location j in cell population. Thus the nucleosome occupancy defined above characterizes the relative probability of a genomic location being covered by a nucleosome in cell population.

We further define the occupancy of a nucleosome (used in Fig. 6C, Fig. S14-16) as the occupancy of its dyad position.

1.7 Poly(dA:dT) content

The poly(dA:dT) content (for example, in Fig. S20) was calculated as follows: for each base pair i in the genome, the local poly(dA:dT) content at i was defined as the poly(dA:dT) content of the 147-bp region s_i centered at i. The poly(dA:dT) content of s_i was subsequently defined as the total length of poly(dA:dT) tracts within s_i divided by 147, where poly(dA:dT) tract was defined as a homopolymer of A or T of length > 3. We chose s_i to be of length 147, because when wrapping the sequence s_i around histone octamer, only poly(dA:dT) tracts contacting the histone core would contribute to the DNA bending energy. In the same spirit, we defined the local strength of 10.5-bp mono-nucleotide periodicity at location i (for example, in Fig. 7) as the mono-nucleotide spectral envelope at period 10.5 bp of the 147-bp sequence s_i centered at i, and the local strength of 10.5-bp di-nucleotide periodicity at location i (for example, in Fig. S18) as the di-nucleotide spectral envelope at period 10.5 bp of the 148-bp sequence s_i centered at i (from $i - 74$ to $i + 73$).

1.8 Additional details

The distribution of p-values in Fig. 2B, Fig. S1B, Fig. S2B, D, Fig. S3B, D, Fig. S17B, D, F, H and the distribution of spectral envelopes at period 10.5 bp in Fig. S12 were smoothed using a Gaussian kernel with bandwidth determined by the Silverman's rule and default settings in the function SmoothKernelDistribution of Mathematica, where the kernel function was specified to account for the domain of the underlying density.

Supplementary References

[1] Peter Bloomfield. *Fourier analysis of time series: an introduction*. John Wiley & Sons, 2004.

[2] Robert H Shumway and David S Stoffer. *Time series analysis and its applications*. Springer Science & Business Media, 2013.

[3] Belinda Phipson and Gordon K Smyth. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology*, 9(1), 2010.

[4] CJ Clopper and Egon S Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, pages 404–413, 1934.

[5] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

[6] Axel Gandy and Georg Hahn. MMCTest–a safe algorithm for implementing multiple monte carlo tests. *Scandinavian Journal of Statistics*, 41(4):1083–1101, 2014.

[7] Daniel Kurz, Horst Lewitschnig, and Jürgen Pilz. Decision-theoretical model for failures which are tackled by countermeasures. *Reliability, IEEE Transactions on*, 63(2):583–592, 2014.

[8] Jun S Song and David E Fisher. Nucleosome positioning in promoters: significance and open questions. *Epigenomics: From Chromatin Biology to Therapeutics*, pages 47–60, 2012.

[9] S Rao Jammalamadaka and Ambar Sengupta. *Topics in circular statistics*, volume 5. World Scientific, 2001.

[10] Timothy J Richmond and Curt A Davey. The structure of DNA in the nucleosome core. *Nature*, 423(6936):145–150, 2003.

[11] Georgette Moyle-Heyrman, Tetiana Zaichuk, Liqun Xi, Quanwei Zhang, Olke C Uhlenbeck, Robert Holmgren, Jonathan Widom, and Ji-Ping Wang. Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning. *Proceedings of the National Academy of Sciences*, 110(50):20158–20163, 2013.

[12] Jonathan Widom. Role of DNA sequence in nucleosome stability and dynamics. *Quarterly reviews of biophysics*, 34(03):269–324, 2001.

[13] B Franklin Pugh. A preoccupied position on nucleosomes. *Nature structural & molecular biology*, 17(8):923–923, 2010.

[14] Noam Kaplan, Timothy R Hughes, Jason D Lieb, Jonathan Widom, and Eran Segal. Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. *Genome Biol*, 11(11):140, 2010.

[15] Kristin Brogaard, Liqun Xi, Ji-Ping Wang, and Jonathan Widom. A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 486(7404):496–501, 2012.

[16] Liqun Xi, Kristin Brogaard, Qingyang Zhang, Bruce Lindsay, Jonathan Widom, and Ji-Ping Wang. A locally convoluted cluster model for nucleosome positioning signals in chemical maps. *Journal of the American Statistical Association*, 109(505):48–62, 2014.

[17] Karolin Luger, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260, 1997.

2 Supplementary Figures

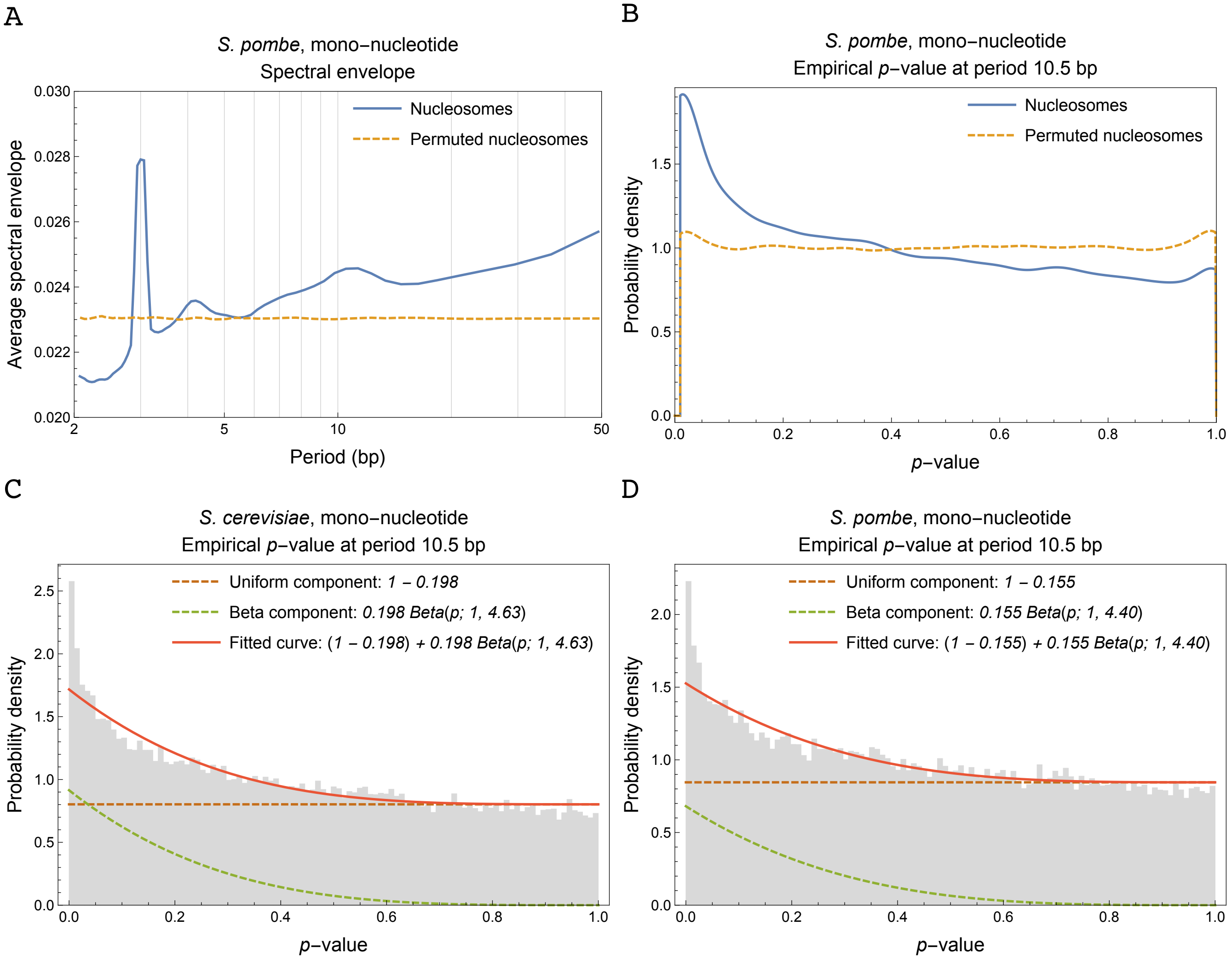

Fig. S1: Only a small fraction of nucleosomal sequences contain significant 10.5-bp periodicity in mono-nucleotides. (A) Average mono-nucleotide spectral envelope of nucleosomal sequences (blue curve) and randomly permuted nucleosomal sequences (yellow curve) in *S. pombe*. (B) Distribution of *p*-values assessing the statistical significance of 10.5-bp periodicity in mono-nucleotides of nucleosomal sequences (blue curve) and randomly permuted nucleosomal sequences (yellow curve) in *S. pombe*. (C) Distribution of *p*-values assessing the statistical significance of 10.5-bp periodicity in mono-nucleotides of *S. cerevisiae* nucleosomal sequences (grey histogram) and the fitted curve using the beta-uniform mixture model (*Supplementary Methods* section 1.1) (solid red curve). Also shown are the fitted uniform component (dashed brown curve) and the fitted beta component (dashed green curve). (D) Same as (C) for *S. pombe*.

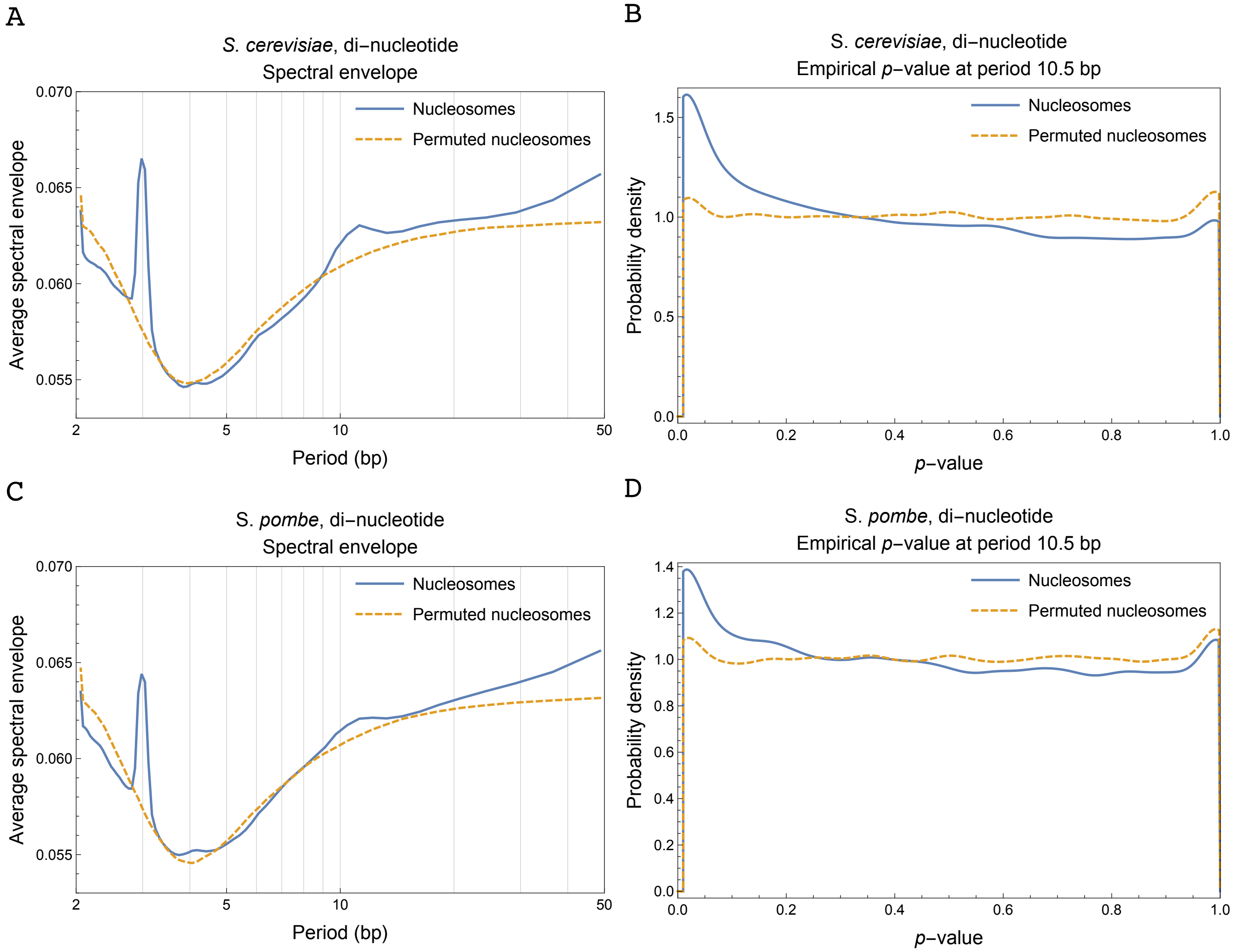

Fig. S2: Nucleosomal sequences in both *S. cerevisiae* and *S. pombe* have enriched 10.5-bp di-nucleotide periodicity compared to randomly permuted sequences. (A) Average di-nucleotide spectral envelope of nucleosomal sequences (blue curve) in *S. cerevisiae* and randomly permuted sequences (yellow curve). (B) Distribution of *p*-values assessing the statistical significance of 10.5-bp di-nucleotide periodicity of nucleosomal sequences in *S. cerevisiae* (blue curve) and randomly permuted sequences (yellow curve). (C,D) Same as (A,B) for *S. pombe*.

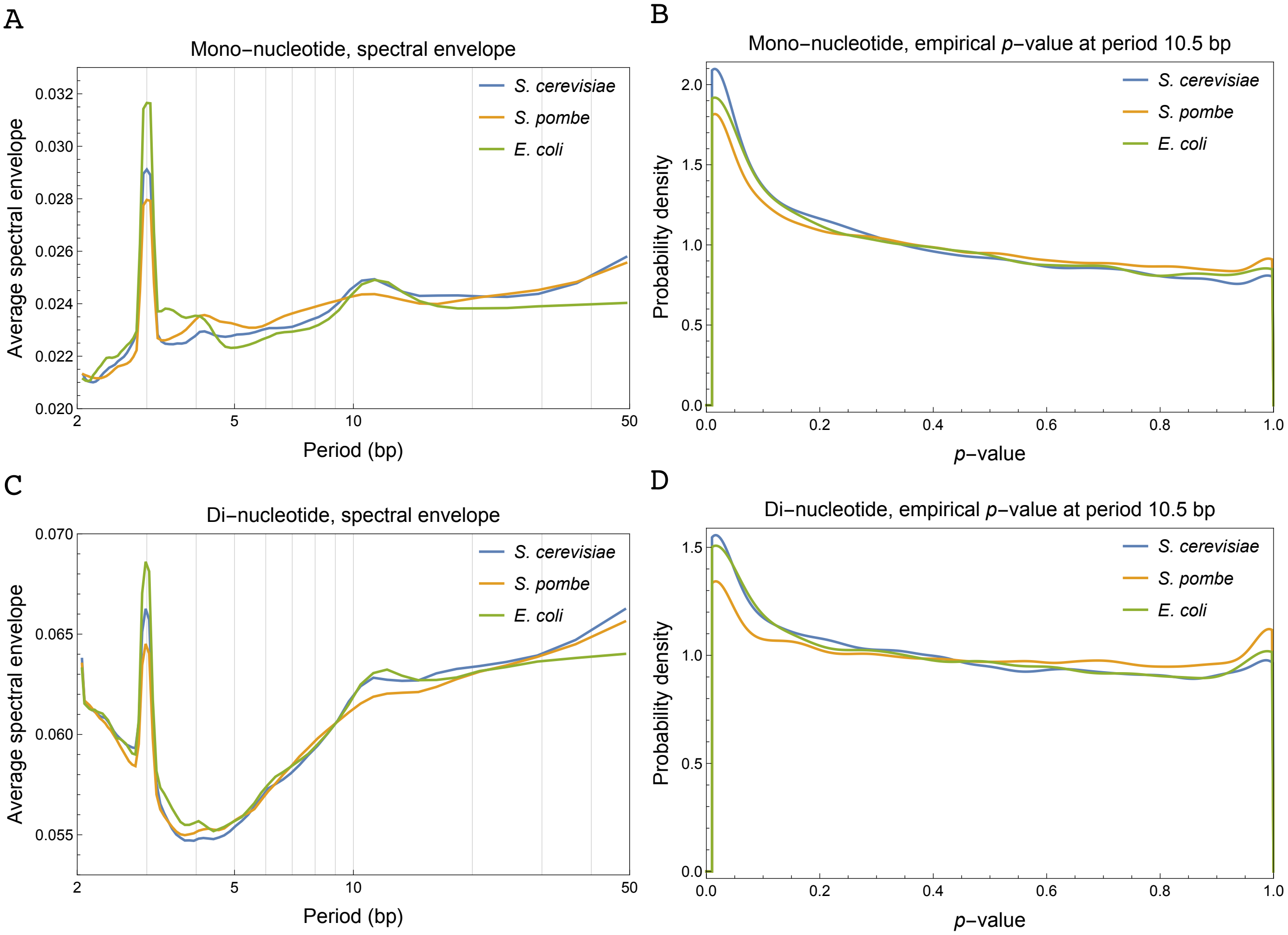

Fig. S3: The strength of 10.5-bp periodicity is comparable among the *E. coli*, *S. cerevisiae*, and *S. pombe* genomes. (A) Average mono-nucleotide spectral envelope of randomly selected 147-bp genomic regions from *E. coli* (25,971 sequences, about the same genome coverage as in *S. cerevisiae* and *S. pombe*), *S. cerevisiae* (67,531 sequences, equal to the number of consensus nucleosomes), and *S. pombe* (75,818 sequences, equal to the number of consensus nucleosomes). (B) Distribution of *p*-values assessing the statistical significance of 10.5-bp mono-nucleotide periodicity of randomly selected 147-bp genomic regions from *E. coli*, *S. cerevisiae*, and *S. pombe*. (C,D) Same as in (A,B) for di-nucleotides.

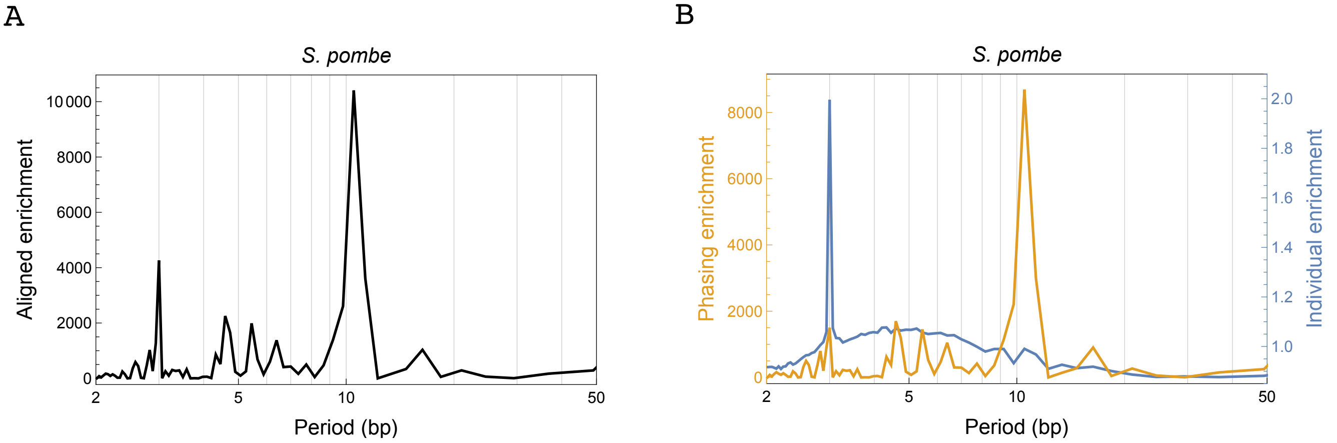

Fig. S4: Spectral decomposition of periodicity in average nucleotide frequency of dyad-aligned nucleosomal sequences in *S. pombe*, where A and T were set to 1 and C and G to 0. (A) The spectrum of aligned enrichment (see main text) for nucleosomal sequences in *S. pombe*. (B) The spectrum of individual enrichment (blue curve) and phasing enrichment (yellow curve) (see main text) in *S. pombe*.

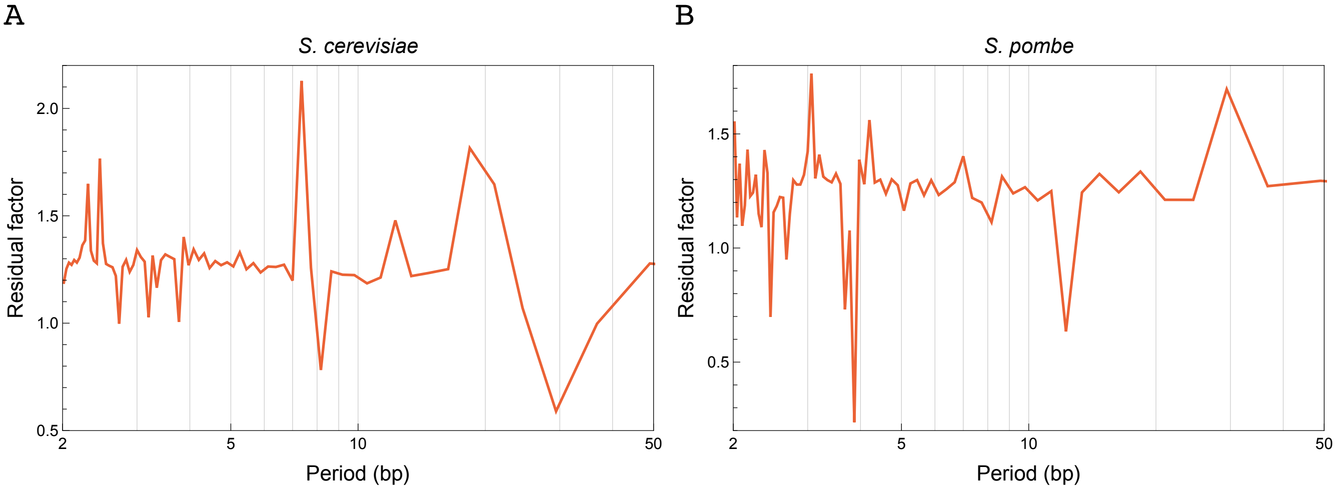

Fig. S5: *Residual factor* (see main text) of the spectral decomposition in *S. cerevisiae* and *S. pombe*, where A and T were set to 1, and C and G to 0. (A) The spectrum of *residual factor* for nucleosomal sequences in *S. cerevisiae*. (B) The spectrum of *residual factor* for nucleosomal sequences in *S. pombe*.

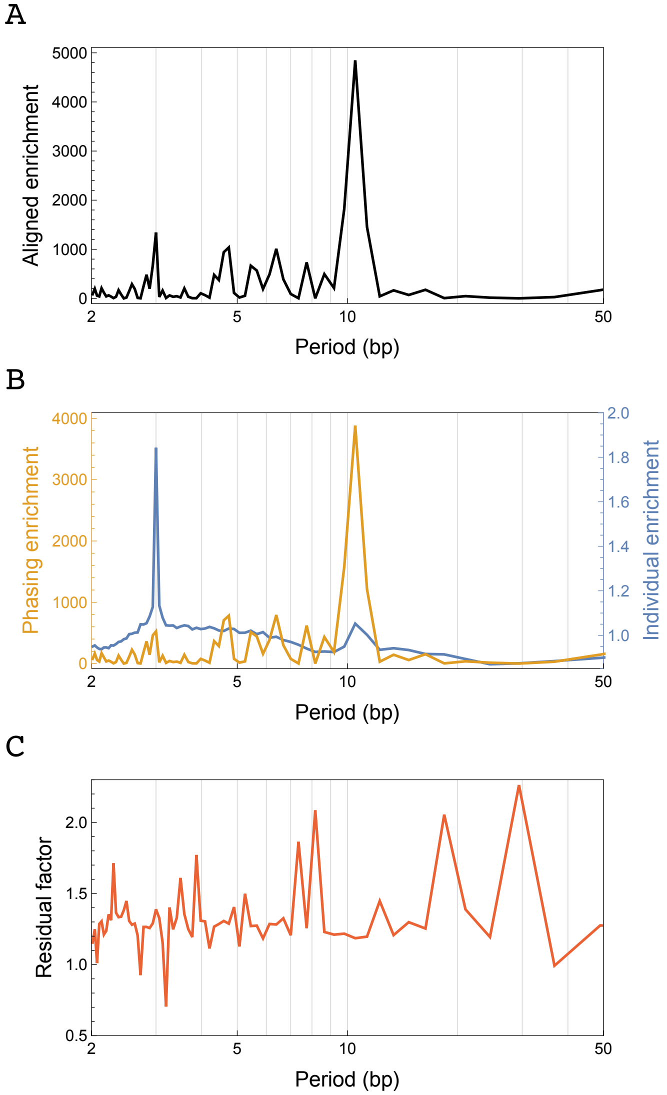

Fig. S6: Subsampling nucleosomal sequences yielded unstable spectrums of the residual factor, whereas the aligned, individual, and phasing enrichments remained stable. We subsampled from nucleosomes in *S. cerevisiae* by taking each nucleosome with probability 0.5, and calculated the spectral decomposition of the subsampled nucleosomal sequences, with A and T set to 1, and C and G to 0. Here we show the result of one instance of subsampling. (A) Spectrum of aligned enrichment of the subsampled nucleosomal sequences. (B) Spectrum of individual enrichment (blue curve) and phasing enrichment (yellow curve) of the subsampled nucleosomal sequences. (C) Spectrum of residual factor of the subsampled nucleosomal sequences. The residual factor has clearly changed from Fig. S5.

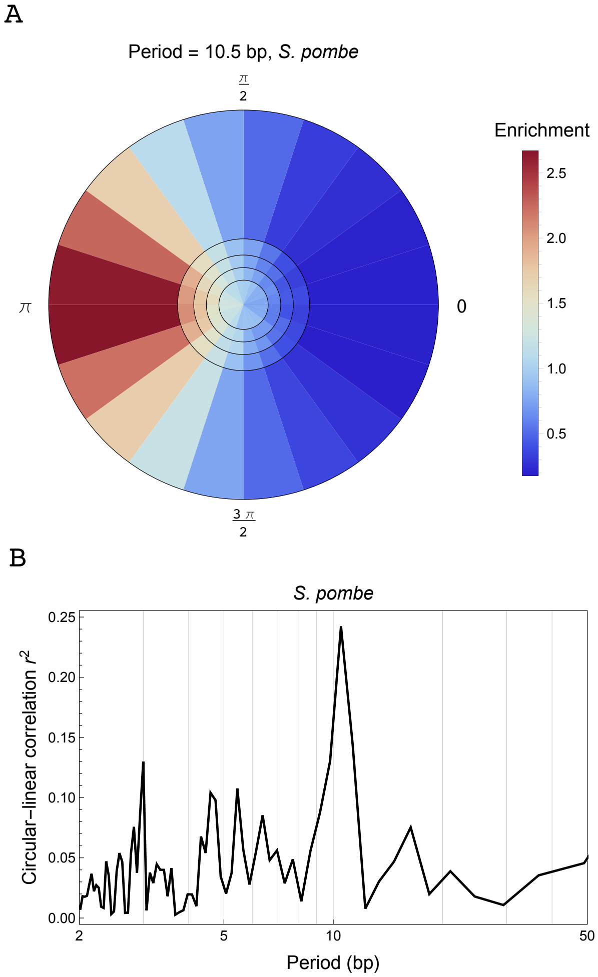

Fig. S7: Phases of nucleosomal sequences in *S. pombe* at 10.5-bp periodicity. For this analysis, A and T were set to 1, and C and G to 0. (A) Phasing of sequences is evident in the Fourier space at period 10.5 bp. Nucleosomes were ranked according to the strength of 10.5-bp periodicity and divided into 5 quintiles (the five circular rings separated by black circles). The phases (with respect to the dyad) of their discrete Fourier transform at period 10.5 bp were then binned into 20 equal intervals (the 20 bins within each ring). Colors indicate the fraction of nucleosomes lying in each bin divided by the expectation under a uniform null distribution (*Supplementary Methods* section 1.4). (B) Circular-linear correlation between the phase and strength of 10.5-bp periodicity (*Supplementary Methods* section 1.4).

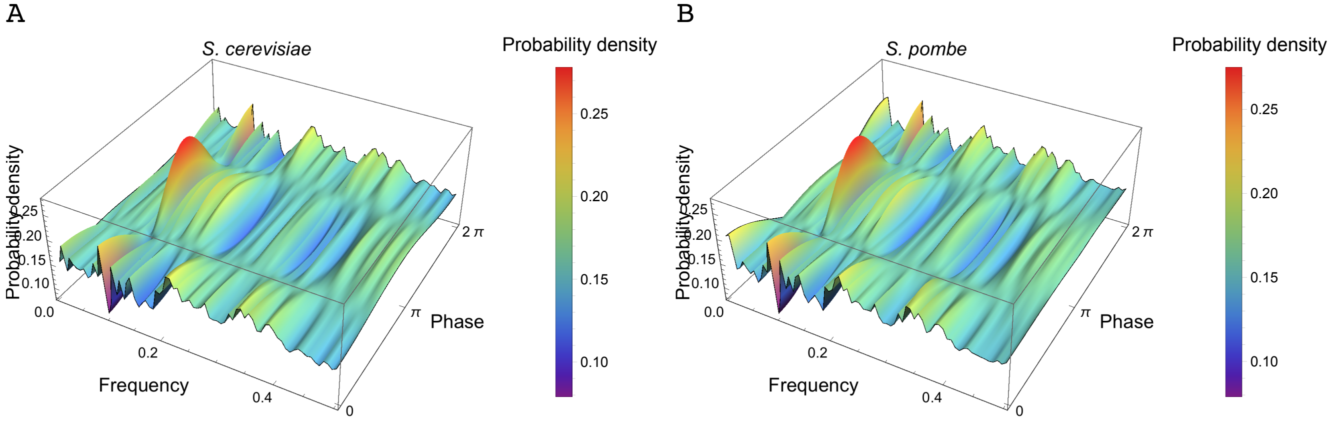

Fig. S8: Phasing among nucleosomal sequences is unique for period 10.5 bp and does not occur at any other periods. (A) Nucleosomal sequences in *S. cerevisiae* were converted to numerical sequences by setting A and T to 1, and C and G to 0. We then calculated the discrete Fourier transform of the numerical sequences and obtained the argument (phase) of the discrete Fourier transform with respect to the dyad at each fundamental frequency (*Supplementary Methods* section 1.4). Figure shows the distribution of these phases, smoothed using a Gaussian kernel with bandwidth determined by the Silverman's rule, and default settings in the function SmoothKernelDistribution of Mathematica. To remove boundary effects, we first augmented the phases to the range $[-2\pi, 4\pi)$ by adding $\theta \pm 2\pi$ to the data for any phase $\theta \in [0, 2\pi)$. We then calculated the kernel density estimate $\hat{f}(\theta)$ of the phase distribution in $[-2\pi, 4\pi)$. The kernel density estimate of the phase distribution in $[0, 2\pi)$ was taken to be $3\hat{f}(\theta)$. (B) Same as in (A), but for *S. pombe*.

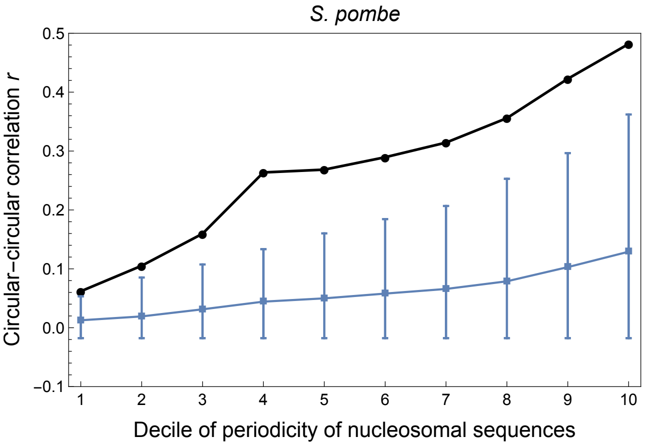

Fig. S9: Redundant nucleosome positions and underlying sequences are highly in phase with each other at period 10.5 bp in *S. pombe*. Figure shows the circular-circular correlation between the phase of nucleotide Fourier transform and the phase of nucleosome location Fourier transform at period 10.5 bp (black curve, *Supplementary Methods* section 1.4). Consensus nucleosomes were first ranked according to the strength of their 10.5-bp nucleotide periodicity and then divided into 10 groups of equal size. The x-axis represents the group index, with group 1 having the smallest strength of 10.5-bp periodicity. Blue curve shows the median correlation coefficient among all fundamental frequencies excluding $1/10.5$, with whiskers showing the range from the 5th percentile to the 95th percentile.

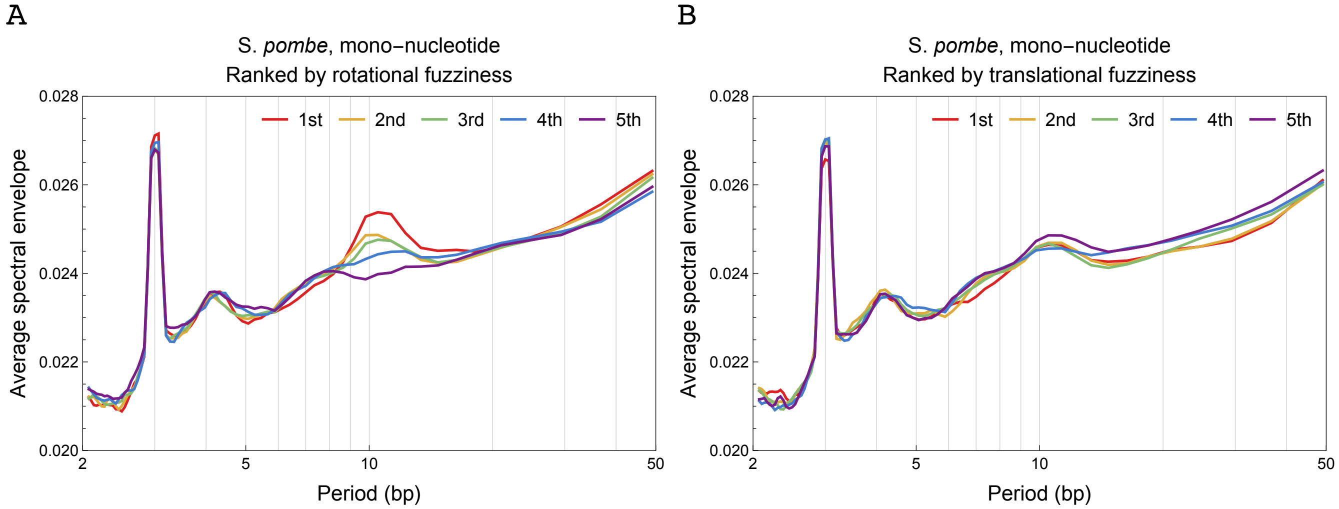

Fig. S10: Spectral envelope of nucleosomes in *S. pombe* grouped by the level of (A) rotational positioning and (B) translational positioning. Nucleosomes in *S. pombe* were ranked from small to large values by rotational fuzziness and translational fuzziness, respectively. The average mono-nucleotide spectral envelope of nucleosomal sequences within each quintile was then plotted, where the 1st quintile contains nucleosomes with the smallest rotational fuzziness and translational fuzziness, respectively.

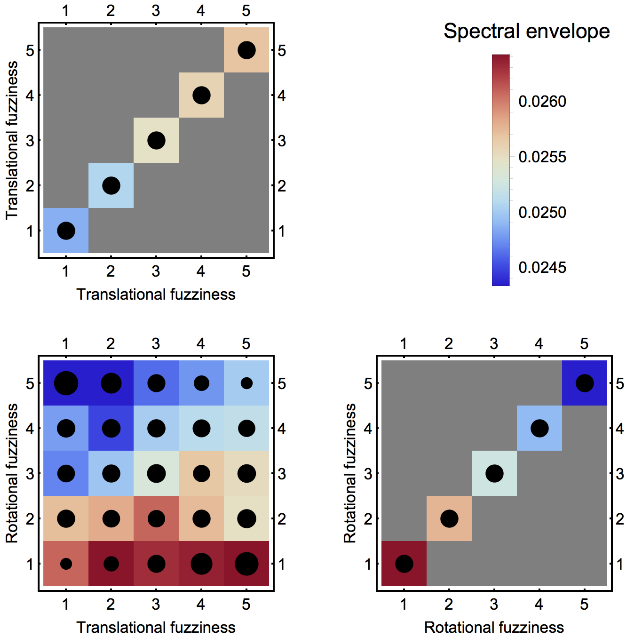

Fig. S11: The slight correlation between translational fuzziness and 10.5-bp periodicity in *S. cerevisiae* (Fig. 6B) results from the anti-correlation between translational and rotational fuzziness. In the top left panel, we ranked the nucleosomes in *S. cerevisiae* by translational fuzziness and divided them into 5 groups of equal size. For example, the (1,1) square represents the quintile with the smallest translational fuzziness. The squares were colored by the average mono-nucleotide spectral envelope at period 10.5 bp within the corresponding groups. The size of the black dot at the center of the squares represents the relative number of nucleosomes within that group. The top left panel shows that the 5 quintiles have the same number of nucleosomes by definition, and there is a slight correlation between 10.5 bp periodicity and translational fuzziness (blue to red color from (1,1) to (5,5) square), consistent with Fig. 6B. The bottom right panel is like the top left panel, but the nucleosomes were ranked by rotational fuzziness instead. The bottom right panel shows a distinct anti-correlation between 10.5-bp periodicity and rotational fuzziness, consistent with Fig. 6A. In the bottom left panel, we ranked the nucleosomes simultaneously by translational and rotational fuzziness, and binned the nucleosomes using the same intervals as in the top left and bottom right panels. For example, the (1,1) square contains nucleosomes that lie in the (1,1) square of the top left panel and the (1,1) square of the bottom right panel simultaneously, i.e. the nucleosomes with the smallest rotational and translational fuzziness. The bottom left panel shows that as rotational fuzziness increases (from 1 to 5), more and more nucleosomes contain small translational fuzziness, since, for example, the size of black dots increases from (1,1) to (1,5) square. We therefore concluded that the observed slight correlation between translational fuzziness and 10.5-bp periodicity (top left panel, Fig. 6B) is an indirect consequence of this anti-correlation between translational and rotational fuzziness.

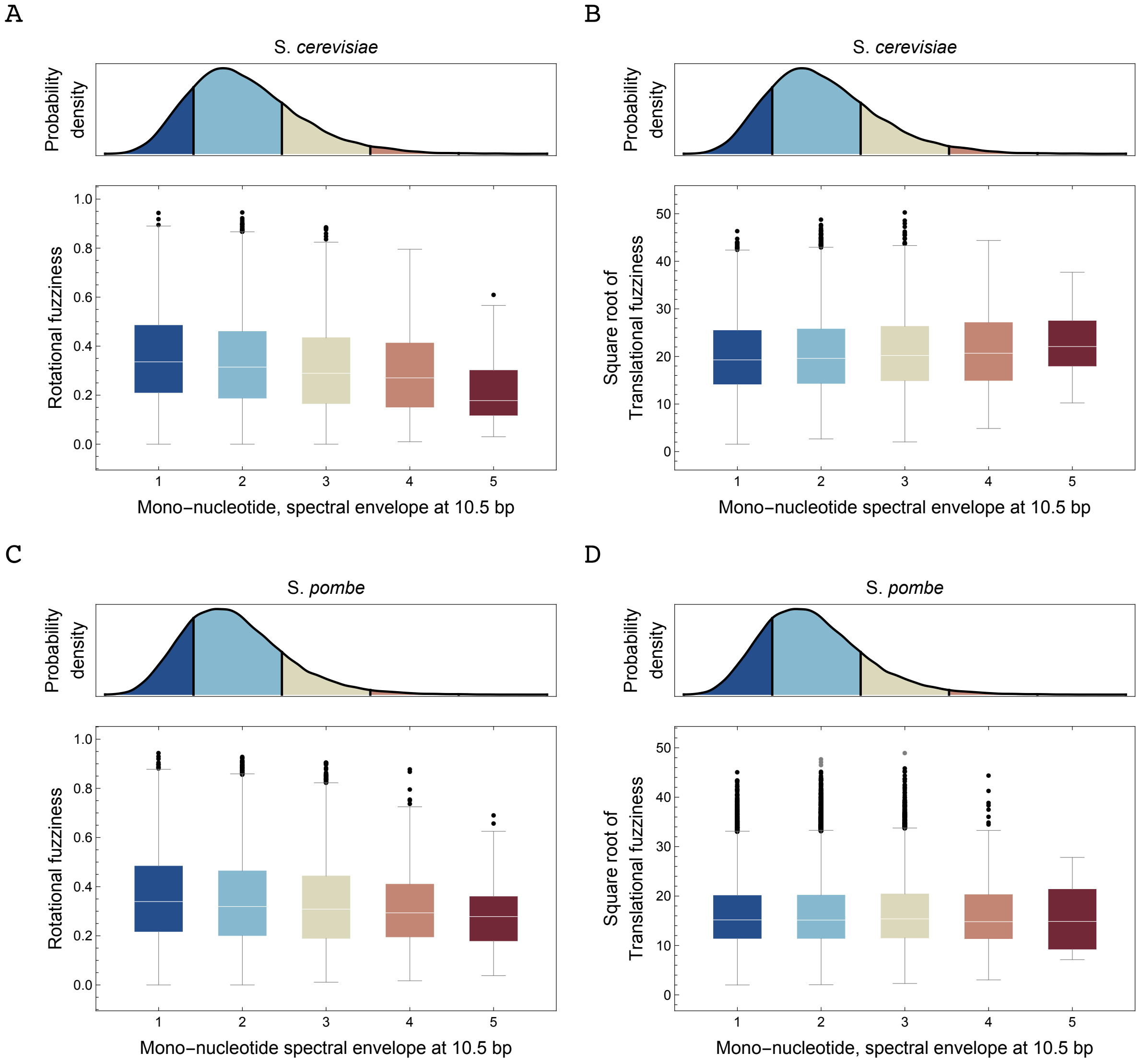

Fig. S12: 10.5-bp periodicity facilitates rotational but not translational positioning. (A) The top panel shows the distribution of the 10.5-bp mono-nucleotide spectral envelope of individual nucleosomes in *S. cerevisiae*. We binned the 10.5-bp spectral envelope into 5 intervals of equal size, shown as different colors in the top panel. The bottom panel is a box plot showing the distribution of rotational fuzziness of individual nucleosomes within each bin. (B) Same as in (A), with the box plot showing the distribution of square root of translational fuzziness. (C) Same as in (A), but for *S. pombe*. (D) Same as in (B), but for *S. pombe*.

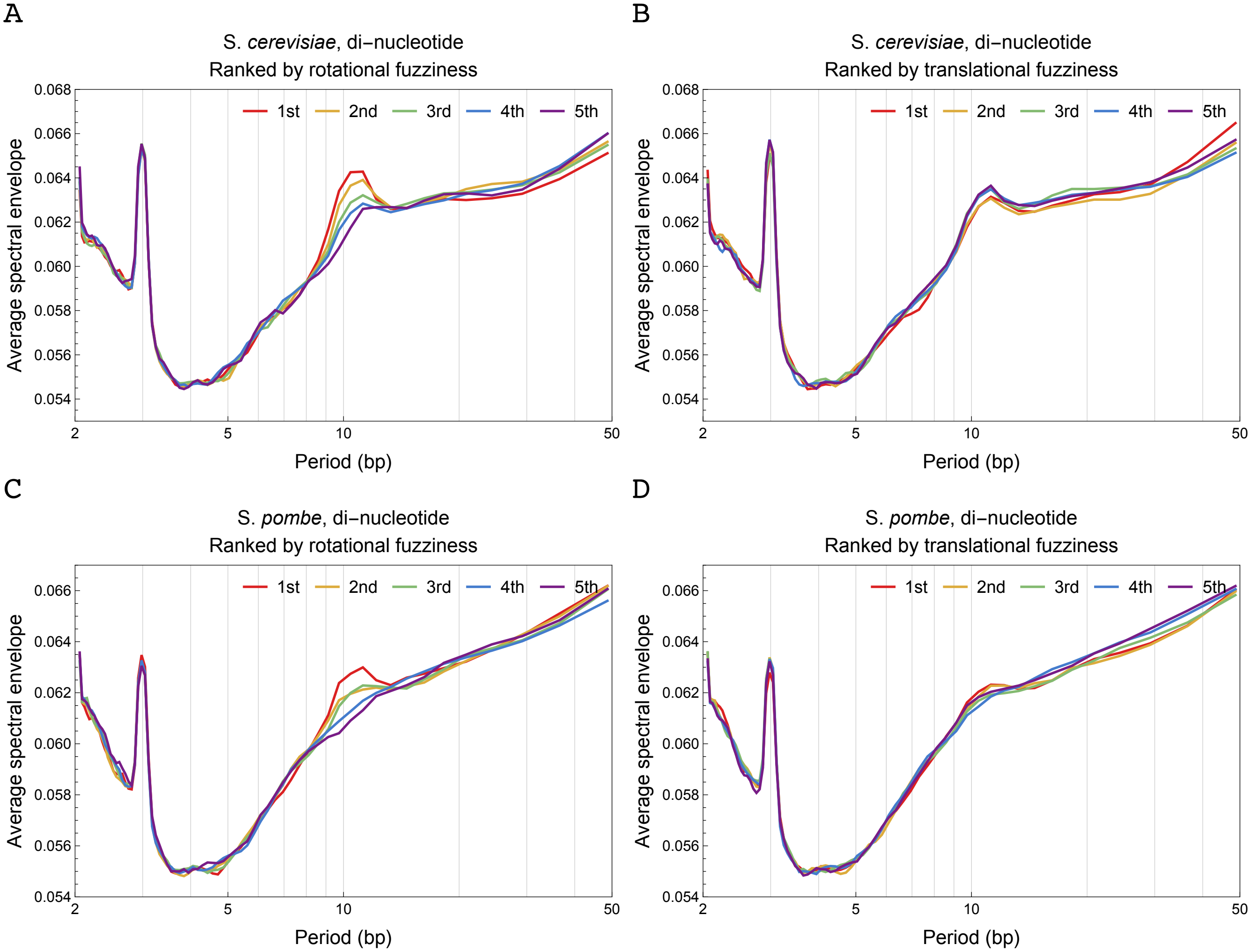

Fig. S13: Di-nucleotide spectral envelope of nucleosomes in *S. cerevisiae* and *S. pombe* grouped by the level of rotational and translational positioning. (A) Nucleosomes in *S. cerevisiae* were ranked from small to large values by rotational fuzziness. Figure shows the average di-nucleotide spectral envelope of nucleosomal sequences within each quintile, where the 1st quintile contains the nucleosomes with the smallest rotational fuzziness. (B) Same as in (A), but with nucleosomes ranked by translational fuzziness. (C) Same as in (A), but for *S. pombe*. (D) Same as in (B), but for *S. pombe*.

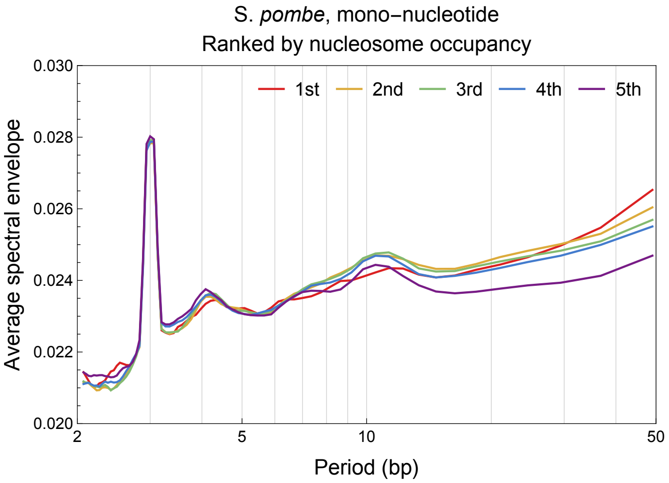

Fig. S14: Spectral envelope of nucleosomes in *S. pombe* grouped by the level of nucleosome occupancy. Nucleosomes in *S. pombe* were ranked from small to large values by nucleosome occupancy. Figure shows the average mono-nucleotide spectral envelope of nucleosomal sequences within each quintile, where the 1st quintile contains the nucleosomes with the smallest nucleosome occupancy.

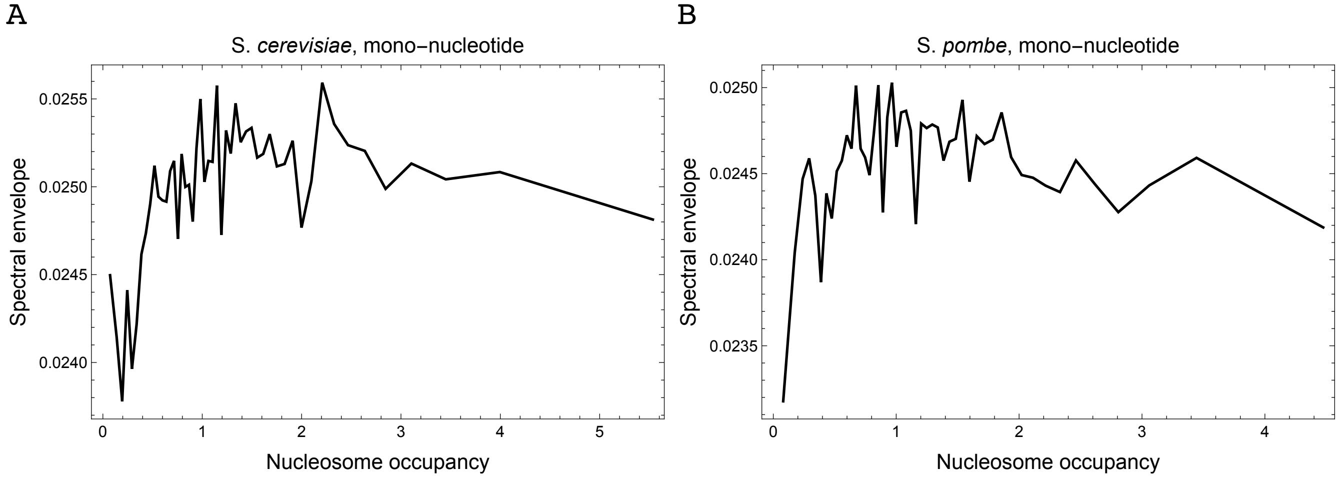

Fig. S15: A finer look at the contribution of 10.5-bp periodicity to nucleosome occupancy in *S. cerevisiae* and *S. pombe*. (A) Nucleosomes in *S. cerevisiae* were ranked from small to large values of nucleosome occupancy, where the occupancy was normalized to genome-wide average. We then divided the nucleosomes into 50 groups of equal size and plotted the average mono-nucleotide spectral envelope at period 10.5 bp against the average nucleosome occupancy within each group. (B) Same as in (A), but for *S. pombe*.

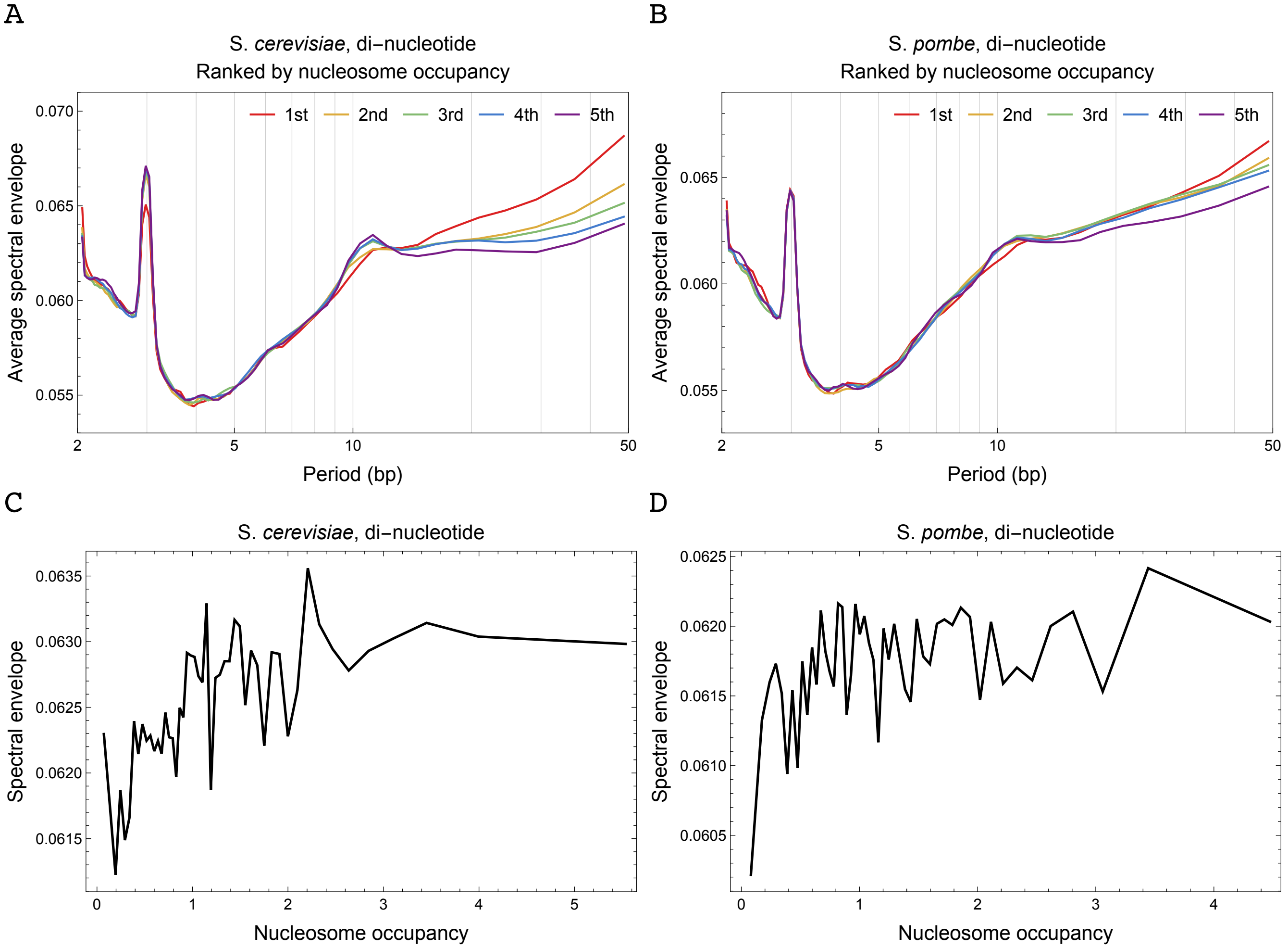

Fig. S16: The contribution of 10.5-bp di-nucleotide periodicity to nucleosome occupancy is only modest. (A) Nucleosomes in *S. cerevisiae* were ranked from small to large values of nucleosome occupancy. Figure shows the average di-nucleotide spectral envelope of nucleosomal sequences within each quintile, where the 1st quintile contains the nucleosomes with the smallest nucleosome occupancy. (B) Same as in (A), but for *S. pombe*. (C) Nucleosomes in *S. cerevisiae* were ranked from small to large values by nucleosome occupancy, where the occupancy is normalized to genome-wide average. We then divided the nucleosomes into 50 groups of equal size and plotted the average di-nucleotide spectral envelope at period 10.5 bp against the average nucleosome occupancy within each group. (D) Same as in (C), but for *S. pombe*.

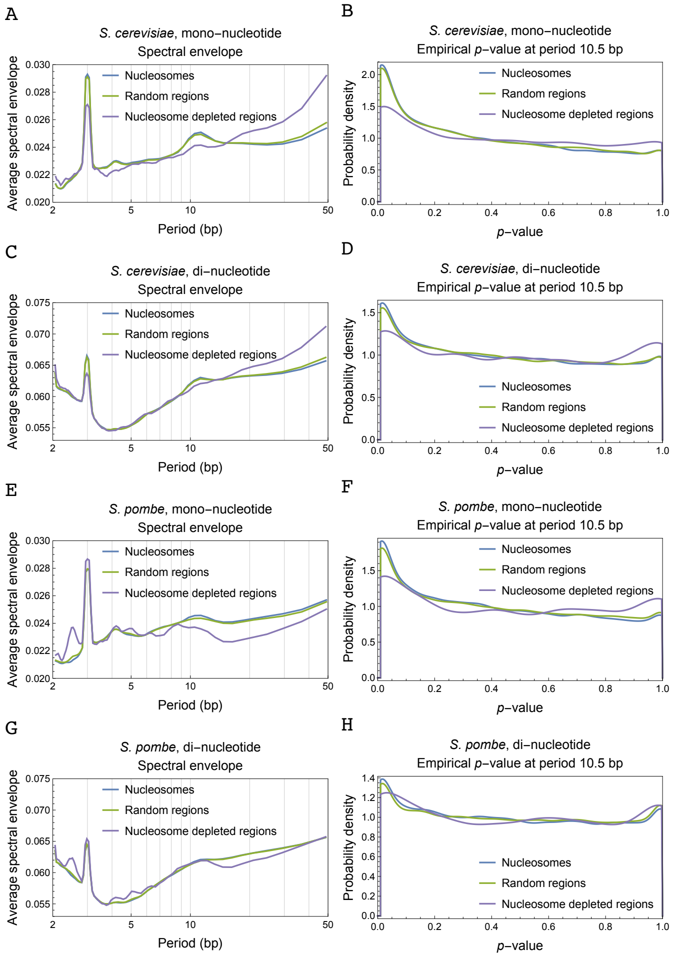

Fig. S17: Randomly selected genomic regions contain a comparable amount of 10.5-bp periodicity as nucleosomal sequences, while nucleosome depleted regions (NDRs) contain much less periodicity. (A) The average mono-nucleotide spectral envelope of 67,531 consensus nucleosomal sequences (blue curve), 67,531 randomly selected 147-bp genomic regions (green curve), and 4,113 linker regions with length greater than 147-bp (NDRs) (purple curve) in *S. cerevisiae*. (B) The distribution of empirical *p*-value assessing the statistical significance of 10.5-bp periodicity in mono-nucleotides for the same groups of sequences as in (A). (C,D) Same as (A,B), but for di-nucleotides. (E) The average mono-nucleotide spectral envelope of 75,818 consensus nucleosomal sequences (blue curve), 75,818 randomly selected 147-bp genomic regions (green curve), and 4,180 linker regions with length greater than 147-bp (NDRs) (purple curve) in *S. pombe*. (F) The distribution of empirical *p*-value assessing the statistical significance of 10.5-bp periodicity in mono-nucleotides for the same groups of sequences as in (E). (G,H) Same as in (E,F), but for di-nucleotides.

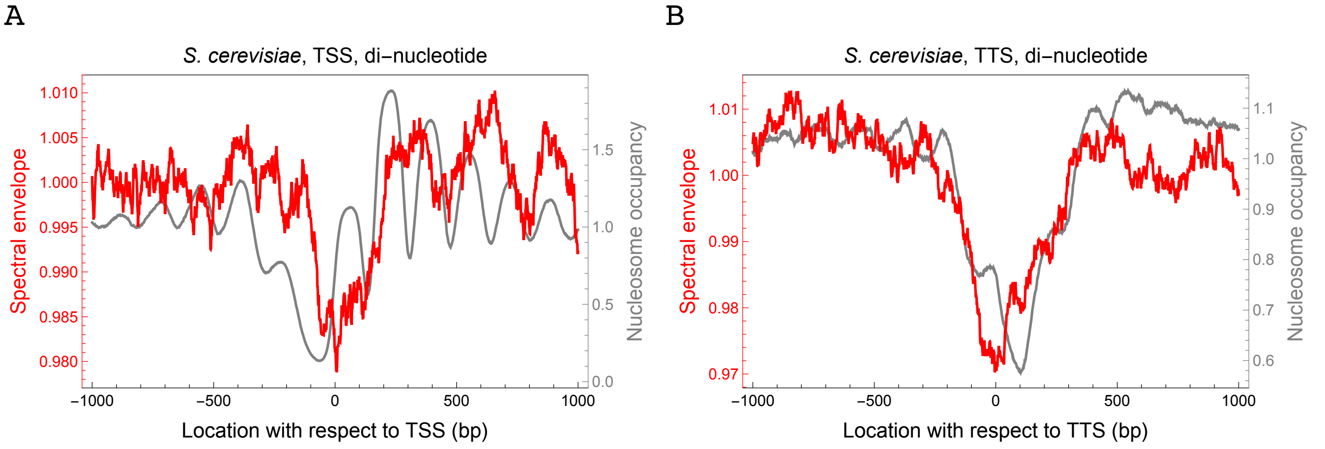

Fig. S18: Nucleosome-depleted regions in *S. cerevisiae* contain reduced 10.5-bp periodicity in di-nucleotides. (A) For each genomic location i, we computed the di-nucleotide spectral envelope at period 10.5 bp of the 148-bp region centered at that location (from $i - 74$ to $i + 73$), normalized to the genome-wide mean. Figure shows the spectral envelope aligned at TSS of 3,005 genes (main text, *Materials and Methods*) in *S. cerevisiae*, normalized to the genome-wide mean (red curve), together with aligned nucleosome occupancy (grey curve). (B) Same as in (A), but aligned at TTS.

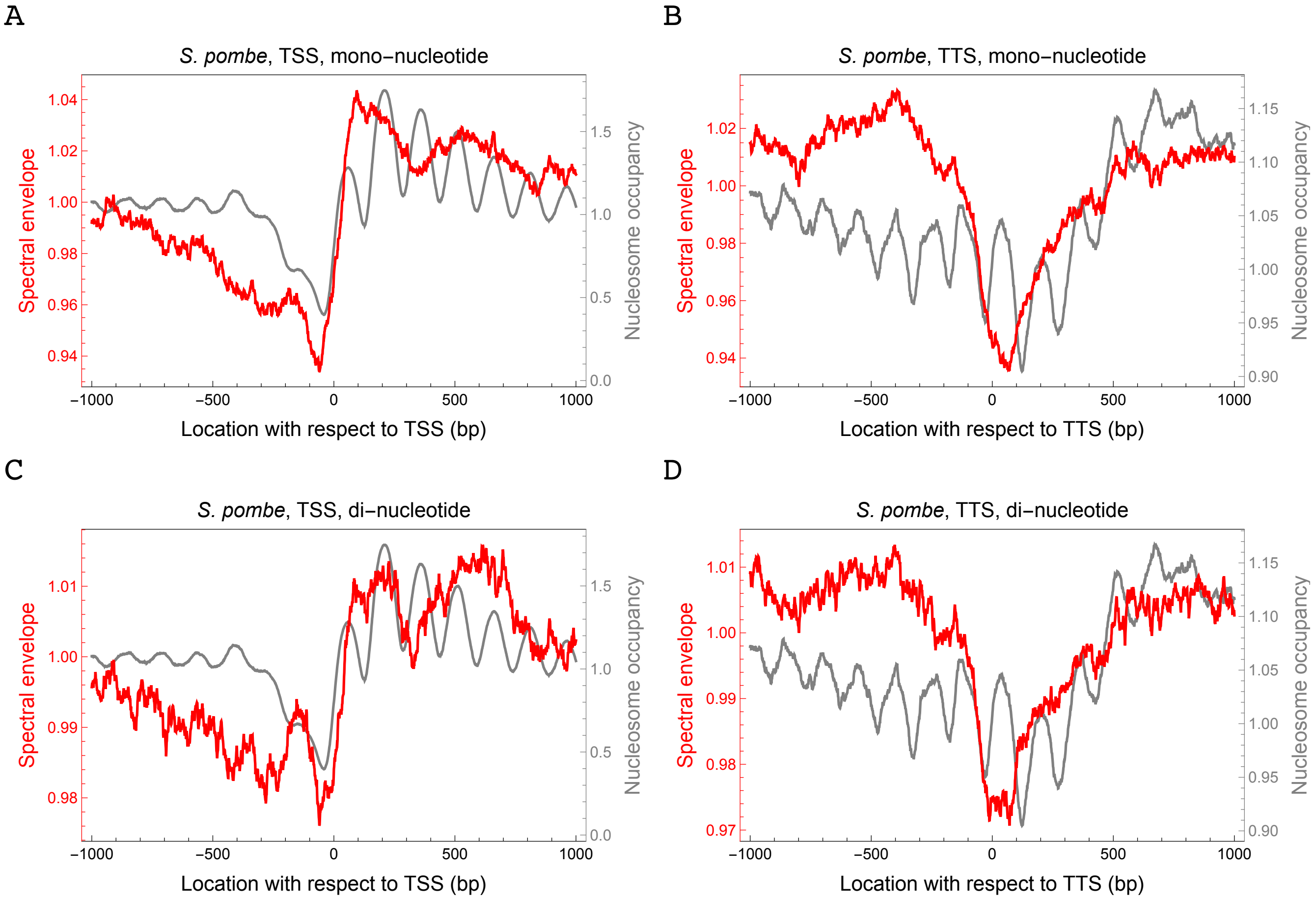

Fig. S19: Nucleosome-depleted regions in *S. pombe* contain reduced 10.5-bp periodicity. (A) At each genomic location i, we computed the mono-nucleotide spectral envelope at period 10.5 bp of the 147-bp region centered at that location, normalized to the genome-wide mean. Figure shows the spectral envelope aligned at TSS of 3,692 genes (main text, *Materials and Methods*) in *S. pombe*, normalized to the genome-wide mean (red curve), together with aligned nucleosome occupancy (grey curve). (B) Same as in (A), but aligned at TTS. (C,D) Same as in (A,B), but for di-nucleotides.

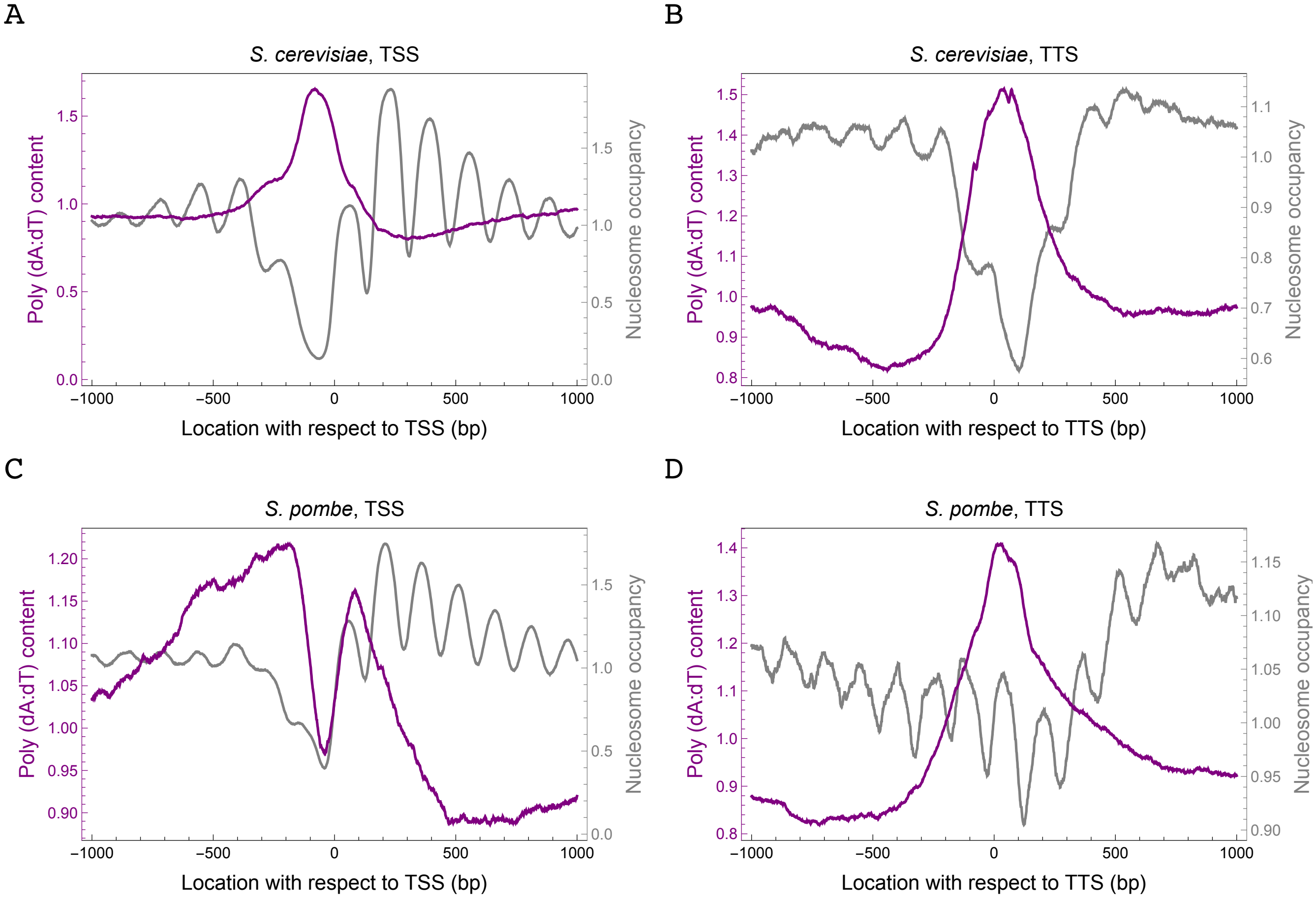

Fig. S20: Poly(dA:dT) contents around TSS and TTS in *S. cerevisiae* and *S. pombe* (*Supplementary Methods* section 1.7). (A) Poly(dA:dT) content (normalized to the genome-wide mean) and nucleosome occupancy (normalized to the genome-wide mean) around TSS in *S. cerevisiae*. (B) Same as in (A), but around TTS. (C,D) Same as in (A,B), but for *S. pombe*.

3 Supplementary Tables

Table S1: Fraction of sequences containing 10.5-bp periodicity in *S. cerevisiae*, *S. pombe* and *E. coli*. We calculated empirical *p*-values characterizing the statistical significance of 10.5-bp periodicity in mono- and di-nucleotides of nucleosomal sequences (nucleosomes), randomly selected genomic regions (random regions), and linker regions of length $>$ 147 bp (nucleosome depleted regions) in both *S. cerevisiae* and *S. pombe* (*Supplementary Methods* section 1.1.2). We also calculated empirical *p*-values characterizing the statistical significance of 10.5-bp periodicity in mono- and di-nucleotides of randomly selected genomic regions (random regions) in *E. coli*. We then fitted the distribution of the empirical *p*-values using a beta-uniform mixture model (*Supplementary Methods* section 1.1.2). The numbers are the estimated mixing coefficient $\hat{\pi}_1$ of the beta component (Equation S2).

		Mono-nucleotide	Di-nucleotide
S. cerevisiae	Nucleosomes	0.198	0.0879
	Random regions	0.182	0.0825
	Nucleosome depleted regions	0.0807	0.0305
S. pombe	Nucleosomes	0.155	0.0438
	Random regions	0.125	0.0268
	Nucleosome depleted regions	0.0540	0.0302
E. coli	Random regions	0.154	0.0694

Table S2: Upper bounds on the fraction of sequences with significant 10.5-bp periodicity at 5% FDR in *S. cerevisiae*, *S. pombe* and *E. coli* (*Supplementary Methods* section 1.1.2).

		Mono-nucleotide	Di-nucleotide
S. cerevisiae	Nucleosomes	0.046	0.032
	Random regions	0.044	0.031
	Nucleosome depleted regions	0.038	0.028
S. pombe	Nucleosomes	0.040	0.027
	Random regions	0.038	0.025
	Nucleosome depleted regions	0.034	0.024
E. coli	Random regions	0.042	0.031

Table S3: Ratio of the value of each factor at 3 and 10.5 bp, respectively, to the corresponding background average, in *S. pombe*.

$1/f$	$A(f)$	$I(f)$	$P(f)$	$R(f)$
3 bp	35.6	2.03	14.8	1.18
10.5 bp	87.0	1.01	85.8	1.01