SUPPORTING INFORMATION

Towards a unified physical model of nucleosome patterns flanking transcription start sites

Wolfram Möbius, Brendan Osberg, Alexander M. Tsankov, Oliver J. Rando, and Ulrich Gerland

I. SOFT-CORE NUCLEOSOME GAS (SNG) – EFFECTIVE INTERACTION BETWEEN PARTICLES

To integrate the spontaneous transient unwrapping of nucleosomal DNA ('nucleosome breathing') into the nucleosome gas model, we give each nucleosome particle two internal degrees of freedom, which correspond to the degree of unwrapping from each end. In its fully wrapped state, a nucleosome within our SNG model occupies a bases on the DNA, in the sense that two neighboring nucleosomes on the DNA begin to constrain each other conformationally when their dyads are a distance a apart on the DNA (we expect to find $a > 147$ bp, since nucleosomes should constrain each other already before their core-particles are immediately adjacent). For two nucleosomes to come closer than a requires that the inward facing DNA segment on at least one of the neighbors changes its conformation into a state with higher free energy than the ground state of the DNA in an isolated nucleosome. This is a complex process on the molecular scale, with many different contributions to the free energy. However, for simplicity, we refer to the free energy cost only as "unwrapping energy". We assume a linear dependence of the unwrapping energy on the number of released bases on one nucleosome end, with

$$\epsilon > 0 \tag{S1}$$

measuring the cost for one base pair in units of the thermal energy $k_{\mathrm{B}}T$.[1] Conceptually similar models have been considered by Chou [1] and Teif *et al.* [2]. For the explicit definition of the model, we allow only odd values,

$$a = 2w + 1 \ , \tag{S2}$$

with a central base at the dyad and w bases to each side. We denote the two internal degrees of freedom as r and l and refer to them as the number of unwrapped bases on the right and left side, respectively. We allow r, l to take on any value in the range

$$0 \leq r,\, l \leq w \ . \tag{S3}$$

The internal free energy of a nucleosome particle is then simply

$$E(r, l) = (r + l)\epsilon \ , \tag{S4}$$

[1] In the following, we use $k_B T = 1$ and implicitly refer to base pairs as length unit where it facilitates readability.

where we assume that there is no coupling between the unwrapping processes at the two ends. We note that some of the above specific choices that we needed to make for the explicit definition of the model do not affect the model behavior in any significant way. In particular, the precise upper bound for the amount of unwrapping is not important, since highly unwrapped states are energetically very unfavorable,

$$\epsilon w \gg 1 \,, \tag{S5}$$

such that they do not contribute significantly to the statistical observables considered below. Also, whether a is chosen even or odd does not matter, since the data does not permit inference of a to single base pair accuracy.

The available nucleosome positioning data cannot distinguish between different internal nucleosome states. In our model, we therefore consider the internal degrees of freedom as equilibrated and calculate the effective nucleosome-nucleosome interaction. To that end, we consider two neighboring nucleosome particles with a given distance Δx between their dyad positions and sum over the Boltzmann weights of all states compatible with this distance, defining the partition function

$$Z(\Delta x) = \sum_{r,l=0}^{w} \Theta(\Delta x - 2w + r + l) e^{-(r+l)\epsilon} \,, \tag{S6}$$

where we use the convention $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ otherwise for the Heaviside function. The effective interaction free energy $v(\Delta x)$ between two neighboring particles is then given by

$$\frac{v(\Delta x)}{k_{\mathrm{B}}T} = -\ln \frac{Z(\Delta x)}{Z(2w+1)} \,, \tag{S7}$$

where the normalization factor $Z(2w+1)$ ensures that $v(\Delta x \geq 2w+1) = 0$. When the unwrapping penalty is significant, i.e. in the relevant limit of Eq. S5, we can obtain a simple approximative form for the interaction $v(\Delta x)$ by extending the sum in Eq. S6 to infinity,

$$\begin{aligned} Z(\Delta x) &\approx \sum_{r,l=0}^{\infty} \Theta(\Delta x - 2w + r + l)\, e^{-(r+l)\epsilon} = \sum_{n=0}^{\infty} \sum_{j=0}^{n} \Theta(\Delta x - 2w + n)\, e^{-n\epsilon} \\ &= \sum_{n=2w-\Delta x+1}^{\infty} (n+1)\, e^{-n\epsilon} = \left(1 - \frac{\partial}{\partial \epsilon}\right) \sum_{n=2w-\Delta x+1}^{\infty} e^{-n\epsilon} \\ &= \left(1 - \frac{\partial}{\partial \epsilon}\right) \frac{e^{-(2w-\Delta x+1)\epsilon}}{1 - e^{-\epsilon}} \,. \end{aligned} \tag{S8}$$

Calculating the derivative and then substituting into Eq. S7 leads to the desired approximation for the interaction free energy,

$$\frac{v(\Delta x)}{k_{\rm B}T} \approx (a-\Delta x)\epsilon - \ln\left[1+(a-\Delta x)\left(1-{\rm e}^{-\epsilon}\right)\right] \quad \text{for} \quad \Delta x \le a = 2w+1 \;, \tag{S9}$$

valid for nucleosome distances Δx not too much smaller than the interaction range $a = 2w+1$. To illuminate the physics of this interaction, it is useful to note that the "self-energy" of a soft-core nucleosome particle is

$$v_0 = -k_{\rm B}T \,\ln\left(\sum_{n=0}^{\infty}{\rm e}^{-n\epsilon}\right)^2 = -2\,k_{\rm B}T\,\ln\frac{1}{1-e^{-\epsilon}} \tag{S10}$$

within the approximation used above (extension of the sum to infinity). The self-energy is the free energy of a single, isolated particle, obtained from the partition function over its internal degrees of freedom. The factor 2 results from the two DNA ends of a nucleosome. The self-energy is also directly related to the normalization factor in Eq. S7 via

$$v_0 = -k_{\rm B}T\,\ln Z(a) \;, \tag{S11}$$

since $Z(a)$ is also the statistical weight for all unwrapping states of two DNA ends, in this case the upstream end of the downstream nucleosome and *vice versa*. We now rewrite Eq. S8 in the form

$$Z(\Delta x) \approx e^{-(a-\Delta x)\epsilon}\cdot\left[\frac{1}{(1-e^{-\epsilon})^2}+\frac{a-\Delta x}{1-e^{-\epsilon}}\right] \;, \tag{S12}$$

which can be interpreted as follows. The first factor is simply the statistical weight (Boltzmann factor) of the minimal unwrapping energy cost required to allow a distance Δx between the two nucleosomes. The second factor contains two terms: The first is equal to $Z(a)$ and describes all unwrapping states of the two DNA ends, starting from one of the ground states, whereas the second term accounts for the degeneracy of the ground state for $\Delta x < a$ and therefore contains only the sum over one DNA unwrapping degree of freedom.

II. SOFT-CORE NUCLEOSOME GAS (SNG) – CALCULATION OF OBSERVABLES

It is long known in statistical physics that the equilibrium properties of so-called "many-body systems" with finite-range interactions in one spatial dimension can be calculated

exactly, see e.g. [3] for a compendium of solved models and [4] for a textbook exposition. The transfer matrix method (see e.g. [5] for a brief pedagogical introduction) is the main tool for discrete systems. These systems consist of lattice sites that can take on different states, with the states of neighboring sites coupled via an energy function (the most prominent example being the Ising model). The method not only permits the exact calculation of thermodynamic observables (such as the free energy or entropy), but also of all statistical observables (such as correlation functions). Our soft-core nucleosome gas model falls within the class of models amenable to this method, as do many related 1D models for collective protein-DNA interactions, see e.g. [2, 6].

Given a system where each lattice site j can take on a finite number S of possible states, the basic idea of the method is to consider the conditional statistical weight $T_{\eta\nu}(j)$ of finding site j in state ν given that site $j-1$ is in state η. This defines an $S \times S$ matrix $T(j)$, the transfer matrix, for each lattice site. Summing over the states of a lattice site j then corresponds to a matrix multiplication, $T(j) \cdot T(j+1)$, and the matrix elements of this product matrix are the statistical weights of finding site $j+1$ in a specified state given that site $j-1$ is in another specified state. Along the same lines, the statistical partition function (i.e., the total statistical weight of all states) and all statistical observables of interest can then be calculated by simple linear algebra. If the system is uniform, i.e. the transfer matrix is identical for all lattice sites, $T(j) \equiv T$, then products of transfer matrices reduce to powers, T^n, in which case it is useful to diagonalize the transfer matrix and switch to its eigenbasis, see below.

For our SNG model, a natural choice of the transfer matrix involves $S = a + 1$ states[2], since each site of the DNA "lattice" can either be outside of the interaction range of any nucleosome (state 0, the "linker state") or be at any one of the $a = 2w + 1$ positions within the interaction footprint of a nucleosome (state 1 through a, the "footprint states"), see Fig. S18 for illustration. For instance, a site in state $w + 1$ sits at the dyad position of a nucleosome, while a site in state 1 ($2w+1$) is at the left-most (right-most) end of a footprint. To obtain an unambiguous definition of the state for each DNA site, we must also consider the situation that the interaction footprints of two nucleosomes overlap, as shown in the

[2] Note that the same physical model can be formulated with different transfer matrices, even with different size S.

bottom panel of Fig. S18. For this situation we adopt the convention that the state of the DNA site is always determined by the nucleosome on the right (with the direction set by the chromosome coordinate x).

Given this definition of the states, the transfer matrix must be chosen to implement the effective nucleosome-nucleosome interaction discussed in the previous section. In the following, we only discuss the case of a uniform system, with site-independent transfer matrix elements $T_{\eta\nu}$, where η and ν each run over all states, i.e. from 0 to a. Explicitly, T takes the form

$$T = \begin{pmatrix} 1 & \mathrm{e}^{\mu} & 0 & 0 & \dots & 0 \\ 0 & \mathrm{e}^{\mu - v(1)} & 1 & 0 & \dots & 0 \\ 0 & \mathrm{e}^{\mu - v(2)} & 0 & 1 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \mathrm{e}^{\mu - v(2w)} & 0 & 0 & \dots & 1 \\ 1 & \mathrm{e}^{\mu} & 0 & 0 & \dots & 0 \end{pmatrix} , \tag{S13}$$

where μ is the chemical potential and the effective interaction potential $v(\Delta x)$ was defined in Eq. (S7). The matrix element $T_{\eta\nu}$ corresponds to the conditional statistical weight of finding a site in state ν given that the previous site (to the left) is in state η. For instance, $T_{0,0} = T_{a,0} = 1$, since there is no interaction between neighboring empty sites or between an empty site and the site to the left in state a (i.e., at the right-most end of a footprint). The entries in the second column correspond to the situation that the right site marks the beginning of a new nucleosome footprint, and hence each element contains a factor e^{μ}, i.e. the statistical weight for adding a nucleosome particle. In addition, there is a factor $\mathrm{e}^{\mu - v(\Delta x)}$ for all states η that imply a footprint overlap with a nucleosome further to the left. Finally, the sequential progression of states, from left to right, within a nucleosome footprint is contained in T via the matrix elements $T_{\nu-1,\nu} = 1$ for $2 \leq \nu \leq a$.

With the help of the transfer matrix (S13), the statistical partition function for a uniform system with N sites is simply the trace of the N-th power of the transfer matrix,

$$Z = \mathrm{Tr}(T^N) \, . \tag{S14}$$

This expression assumes periodic boundary conditions, i.e. strictly speaking it would apply only for a circular chromosome, but as long as N is large, all statistical observables of interest here are insensitive to the boundary conditions. In particular, the statistical distribution

functions deep inside a long linear chromosome are identical to those on a long circular chromosome. The calculation of the statistical distribution functions is based on the fact that the probability of finding nucleosomes at specified lattice sites can be expressed as the ratio of two partition functions, a constrained partition function and the unconstrained partition function (S14). A constrained partition function is obtained by fixing the states of the lattice sites of interest and summing over the states of the remaining sites. Within the transfer matrix approach this is implemented by inserting projection matrices at the appropriate positions within the product of transfer matrices. For instance, the constrained partition function $Z|_0$ that keeps one particle at a fixed position is

$$Z|_0 = \mathrm{Tr}(F \cdot T^N) \,, \tag{S15}$$

where the projection matrix F with elements $F_{\eta\nu} = \delta_{\eta\nu}\delta_{\eta 1}$ enforces state 1 at one lattice site (it is irrelevant which one, since we consider a uniform system with periodic boundary conditions). In the limit of large N, the ratio of the constrained and full partition functions yields the average nucleosome density,

$$\overline{\rho} = \lim_{N\to\infty} \frac{Z|_0}{Z} \,, \tag{S16}$$

i.e., the average probability of finding a nucleosome dyad[3] at a lattice site ($0 < \overline{\rho} < 1$). Similarly, we can introduce a partition function where two particles are kept at fixed positions with a distance x,

$$Z|_{0,x} = \mathrm{Tr}(F \cdot T^x \cdot F \cdot T^{N-x}) \tag{S17}$$

to obtain the particle density $\rho(x|0)$ at position x given a positioned particle at site 0 via

$$\rho(x|0) = \lim_{N\to\infty} \frac{Z|_{x,0}}{Z|_0} \,. \tag{S18}$$

Transforming the matrices F and T to the eigenbasis of T makes the limit $N \to \infty$ analytically accessible. We denote matrices written in the eigenbasis of T by a hat ($\hat{T}$ then is diagonal), and we define the eigenvalues λ_ν of T in decreasing order of absolute value, i.e.,

[3] Note that due to our convention for overlapping nucleosome footprints, where the state of a DNA site is always the smaller of two possible states, we have to use state 1 for "counting" particles, since this is the only state that necessarily occurs for each particle. However, we consider the dyad position (w bases further to the right from a state 1 position) as the location of a nucleosome.

$|\lambda_0| > |\lambda_1| \geq |\lambda_2| \geq \ldots$, and $\hat{T}_{00} = \lambda_0$, $\hat{T}_{11} = \lambda_1$, ... For large systems, the largest (absolute) eigenvalue, λ_0, dominates the partition function. Since the trace is invariant under a change of basis, one obtains the average nucleosome density

$$\overline{\rho} = \lim_{N\to\infty} \frac{\mathrm{Tr}(\hat{F}\cdot\hat{T}^N)}{\mathrm{Tr}(\hat{T}^N)} = \lim_{N\to\infty} \frac{\lambda_0^N \sum_\kappa \hat{F}_{\kappa\kappa}\left(\frac{\lambda_\kappa}{\lambda_0}\right)^N}{\lambda_0^N \sum_\nu \left(\frac{\lambda_\nu}{\lambda_0}\right)^N} = \hat{F}_{00} \tag{S19}$$

and the local nucleosome density (in the vicinity of a well-positioned nucleosome)

$$\begin{aligned}
\rho(x|0) &= \lim_{N\to\infty} \frac{\mathrm{Tr}(\hat{F}\cdot\hat{T}^x\cdot\hat{F}\cdot\hat{T}^{N-x})}{\mathrm{Tr}(\hat{F}\cdot\hat{T}^N)} \\
&= \lim_{N\to\infty} \frac{\lambda_0^N \sum_{\kappa,\sigma} \hat{F}_{\sigma\kappa}\left(\frac{\lambda_\kappa}{\lambda_0}\right)^x \hat{F}_{\kappa\sigma}\left(\frac{\lambda_\sigma}{\lambda_0}\right)^{N-x}}{\lambda_0^N \sum_\nu \hat{F}_{\nu\nu}\left(\frac{\lambda_\nu}{\lambda_0}\right)^N} \\
&= \frac{1}{\hat{F}_{00}} \sum_\kappa \hat{F}_{0\kappa}\hat{F}_{\kappa 0}\left(\frac{\lambda_\kappa}{\lambda_0}\right)^x
\end{aligned} \tag{S20}$$

that we use to calculate the nucleosome patterns of the SNG model shown in Fig. 2 of the main paper and in the supplementary figures below.

We can also use the transfer matrix to calculate the distribution of distances Δx between neighboring nucleosomes, $p_{nn}(\Delta x)$, by introducing an additional projection operator $G_{\eta\nu} = \delta_{\eta\nu} - F_{\eta\nu}$, which ensures that there is *no* particle located at a site. Repeated use of this projection operator guarantees that no particle is bound in between two specified particles, leading to

$$p_{nn}(\Delta x) = \frac{\sum_{\nu\kappa} \hat{F}_{0\nu}\left([\hat{T}\hat{G}]^{\Delta x-1}\hat{T}\right)_{\nu\kappa}\hat{F}_{\kappa 0}}{\lambda_0^{\Delta x}\hat{F}_{00}}\,. \tag{S21}$$

Finally, we consider a more microscopic observable of the SNG model: the effective DNA footprint size a_{eff}. This quantity captures the average reduction of the footprint a due to DNA unwrapping. Its definition,

$$a_{eff} = a - \langle r + l\rangle\ , \tag{S22}$$

includes both spontaneous DNA breathing (which occurs also at low nucleosome density), as well as pressure-induced unwrapping (which is relevant only at high densities), i.e. the average over the unwrapping lengths r and l (introduced in the previous section) is not a single-nucleosome property but is an effect which must be considered in the context of the

nucleosome gas. We can calculate this average by first summing over all unwrapping states compatible with a fixed spacing Δx between neighboring nucleosomes, and then summing over Δx weighted by its distribution $p_{nn}(\Delta x)$. We find

$$a_{eff} = a + \sum_{\Delta x} p_{nn}(\Delta x) \frac{\partial}{\partial \epsilon} \ln Z(\Delta x) \,, \tag{S23}$$

where $Z(\Delta x)$ is the partition function defined in Eq. S6. This expression is plotted in Fig. S15.

III. DATA PROCESSING

We obtained nucleosome positioning data for 12 yeast species from Tsankov *et al.* [7] and data for 3 different MNase digestion levels in *S. cerevisiae* from Weiner *et al.* [8]. The data from Weiner *et al.* [8] was used only for our study of the effect of digestion levels on our estimate of the effective nucleosome width b, see section VI and Fig. S19. The data was processed in a similar manner as in Ref. [7], with slight modifications. Briefly, mono-nucleosomal DNA was isolated from cells grown to mid-log phase and sequenced using single-end Illumina technology with 36 bp reads. We then used BLAT [9] to map sequenced reads from each species to the corresponding reference genome, keeping only reads that mapped to a unique location, allowing for up to 4 mismatches, and considering only sites where no ambiguous reads were mapped to. To merge the reads from the Watson and Crick strands, we shifted the read start locations by half of the mean DNA fragment length (estimated by cross-correlating the reads from both strands [7]) in the direction towards the nucleosome center and took the sum to represent the total read density.

Using the genomic nucleosome read locations, we inferred nucleosome positions and searched for the 5' NFR with the (downstream) flanking +1 nucleosome for each gene in all the species as described in Ref. [7]. We aligned the read density relative to the +1 nucleosome at all genes where a NFR was identified, and averaged across genes, resulting in $\Omega(x)$, the gene-averaged read density at distance x downstream from the +1 nucleosome for a given species.

Similarly, we determined averages over genes longer than 2000 or 3000 bps and genes with 20% lowest and 20 % highest expression level (considering genes with reported expression level only). Absolute gene expression levels were taken from Ref. [7]. Gene lengths were

obtained using the species annotations used in Ref. [7].

IV. OPTIMIZING PARAMETERS OF THE NUCLEOSOME GAS MODELS

To compare the observed read densities to our theoretical models and to estimate our model parameters, we first computed the theoretical nucleosome patterns over a wide range of biologically reasonable parameters for both the HNG and SNG models. Subsequently, we determined the optimal parameters, in each case, by minimizing the mean-squared deviation per data point between data and model.

In the case of the SNG model, we employed the transfer matrix as explained above to compute the average density $\overline{\rho}$ for a given chemical potential μ (Eq. S19) and the density pattern close to a boundary $\rho(x|0)$ (Eq. S20). Note that although we consider the grand-canonical ensemble here, we may equivalently use the average density $\overline{\rho}$ as a system parameter, given the assumption of a large system size.

For the HNG model (with the average density $\overline{\rho}$ and particle width b as parameters), we used the known analytical result for a continuum semi-infinite system,

$$\rho(x|0) = \delta(x) + \sum_{k=1}^{\infty} \frac{\left(\frac{x}{b} - k\right)^{k-1} \Theta\left(\frac{x}{b} - k\right)}{b \cdot (k-1)!} \left(\frac{\overline{\rho} b}{1 - \overline{\rho} b}\right)^{k} e^{-\frac{x/b-k}{1/\overline{\rho} b - 1}} , \tag{S24}$$

as previously used in Ref. [10].

When optimizing parameters, both the finite width of the density peak around the position of the +1 nucleosome and the unknown sequencing depth had to be taken into account. As explained in the main text, the finite width of the +1 nucleosome peak may originate from a localized attractive region in the effective free energy landscape $u(x)$, restricting the +1 nucleosome to a finite interval. Furthermore, experimental errors could contribute to the finite width of the peak. For simplicity, we convolved the model with the shape of the peak before optimizing parameters. Formally, the model's nucleosome density $\tilde{\rho}(x|0)$ thus depends on the species considered: $\tilde{\rho}(x|0) = \sum_{y=c_1}^{\min(c_2,x)} \rho(x-y|0) \cdot \Omega(y) / \left[\sum_{z=c_1}^{c_2} \Omega(z)\right]$. As convolution range we chose c_1 and c_2 to be the positions where the peak falls to a quarter of its maximum value.

The unknown sequencing depth can be treated as an additional parameter in the optimization process [10]. With α denoting a (species-dependent) normalization factor related

to the sequencing depth, the mean-squared deviation per data point reads

$$\hat{\delta}^2 = \frac{1}{l_2 - l_1 + 1} \sum_{x=l_1}^{l_2} \left[\Omega(x) - \alpha \cdot \tilde{\rho}(x|0)\right]^2 , \tag{S25}$$

from which α can be determined analytically by minimizing $\hat{\delta}^2$ with respect to α, resulting in $\alpha = \left(\sum_{x=l_1}^{l_2} \tilde{\rho}(x|0) \cdot \Omega(x)\right) / \left(\sum_{z=l_1}^{l_2} \tilde{\rho}(z|0)^2\right)$. In the following, we refer to α simply as 'the sequencing depth'. Note, however, that according to its definition, α depends on the model and its parameters. For the data reported in the main manuscript, the comparison between model and data was limited to a window $[l_1, l_2] = [200, 1900]$ (however see below for a robustness analysis which includes an alternative choice of fitting range).

Finally, we optimized parameters by minimizing $\hat{\delta}^2$ over the range of model parameters listed in Table S1, for each species independently. We use $\delta^2 = \hat{\delta}^2/\alpha^2$ to report the deviation between model and data at the level of nucleosome densities (instead of read densities) in Fig. 4 and Fig. S13, in order to make the deviation values more comparable between different species. For all cases reported in the main paper, the parameter values and δ^2 deviations obtained by our optimization procedure are listed in Tables S2 and S3.

In the case of the HNG model, we first performed fits with the nucleosome width kept fixed at $b = 147$ bp, as described in the main text. By inspection of the fits shown in Fig. S1A to S12A and the deviations plotted in Fig. S13, it is clear that the two prominent features of the nucleosome patterns, their period and the shape of the decaying amplitude, cannot be simultaneously accounted for in our different species by the HNG model with fixed nucleosome width. This is apparent, in particular, for the species *K. waltii*, *K. lactis*, and *D. hansenii*. We note that the disagreement between HNG model and data is much more significant than the uncertainty in the data itself, as estimated e.g. by considering the average pattern from different subset of genes from the same species (Fig. S1B to S12B) or by comparing independent experiments for the same species (compare Fig. S1A and S19A for different *S. cerevisiae* data sets). This justifies our conclusion, drawn in the main text, that the hard-core nucleosome gas must be ruled out as a unified physical model.

In the case of the unified SNG model, parameters were optimized for all species except *K. lactis* simultaneously (see main text). To that end, we minimized $\sum_i \omega_i \cdot \hat{\delta}_i^2$ where the sum is taken over the species and the weights $\omega_i = 1/\alpha_{\mathrm{SNG},i}^2$ are introduced to account for varying sequencing depth between species, where $\alpha_{\mathrm{SNG},i}$ is the sequencing depth of species i estimated from the independent SNG fit to that species. In the unified SNG model, the

free energy cost ϵ and maximum wrapping length w are uniform across species, while the chemical potential μ is species-dependent.

To investigate how well the unified SNG model describes the data from *K. lactis*, we optimized the chemical potential μ while the parameters ϵ and w were set to the values obtained from the simultaneous optimization over the 11 other yeast species (see also Table S1).

To test for robustness of our fitting procedure, we repeated the fits for the SNG model also for a different fitting range (starting already at 100 bp from the +1 nucleosome), for a different convolution range (the range over which the peak falls to half of its maximum value), and also for the combination of these two modifications. The results are tabulated in Tables S4, S5, and S6. While these modifications lead to slight changes in the optimal fit parameters, our conclusions are not affected by these details. For the main paper, we have chosen the combination of settings (fit start at 200 bp and convolution up to a quarter of the peak height) that leads to the most conservative result, in the sense that it produces the most modest improvement from HNG to SNG model in the deviation δ^2 between experimental and theoretical pattern.

V. ACTIVE SNG MODEL

We addressed the effect of remodelers shaping the density pattern in addition to the interaction between neighboring particles using kinetic Monte Carlo simulations. Specifically, we considered a fixed number of particles on a periodic lattice with about 8000 sites[4]. Particles were allowed to slide to adjacent lattice sites at rate $k_s \cdot \min\left[1, \exp(-\Delta V/kT)\right]$ where $\Delta V = V_{\text{new}} - V_{\text{old}}$ is the change in total energy that results from the given transition. The total energy V is given by the sum of the interaction potential $v(\Delta x)$ between all nearest-neighbor pairs of particles in the system (described above, parameters $\epsilon = 0.1525\, k_B T/\text{bp}$ and $a = 2w+1 = 167\,\text{bp}$ as estimated for the unified SNG model). A configuration with two particles occupying the same lattice site is considered to have infinite energy, prohibiting this configuration.

In addition to this passive sliding process, remodelers randomly bind to two neighboring nucleosomes at rate k_{RM} and reduce the distance by shifting one nucleosome towards the

[4] The actual system size was increased to the next integer multiple of the average spacing to yield the desired average particle density.

other by one base pair, but only for distances up to $4w$. See Fig. 5A for an illustration.

The kinetics of the system was simulated employing the Gillespie algorithm [11]. Based on random initial conditions and an estimated equilibration time for the local nucleosome density of $t_{eq} = (2k_s\overline{\rho}^2)^{-1}$, the simulation was run for a time $t_{sim} = 100\, t_{eq}$. Two examples for particle trajectories are displayed in Fig. S16.

To determine the nucleosome density close to a wall, we averaged over time (for $t \geq t_{eq}$) and 100 independent realizations. Furthermore, we averaged over reference particles, i.e., no particles were fixed in the simulation, but each particle was considered as a boundary particle for determining $\rho(x|0)$.

VI. EXCLUDING MNASE DIGESTION ARTEFACTS IN THE ESTIMATION OF EFFECTIVE NUCLEOSOME WIDTHS

Our motivation for introducing nucleosome breathing and the SNG model to our analysis was the correlation, obtained within the HNG model, between the best-fit nucleosome width b and the average repeat length $1/\overline{\rho}$, see Fig. 2B. A possible concern is that this correlation is caused by experimental artefacts related to the MNase digestion step in the experimental protocol and the fact that we have only single-end sequencing data. For instance, although the digestion level in each species was performed according to the same criterion (digestion until the trinucleosome band was only barely visible [7]), there may still be a slight variation in MNase digestion level from species to species. Therefore one may ask: Can we distinguish between true nucleosome breathing effects and MNase digestion artefacts in our analysis of the data?

An essential qualitative difference between true nucleosome breathing and MNase digestion artifacts is the following: MNase digestion should not affect the nucleosome dyad positions on the DNA but only blur our observation of these positions, since MNase digestion leads to a variable distance between nucleosomal DNA end and dyad. In contrast, true breathing has a long-range effect on the statistics of nucleosome dyad positions, as described by the soft-core nucleosome gas model. In this work, we obtain an effective nucleosome size by fitting the long-range nucleosome pattern and not by a local measure of nucleosome size. Therefore it should be a robust measure. To test this, we performed two additional analyses below.

One important point to consider is that the experimental protocol included a selection step for mononucleosome-sized DNA fragments. However, this was not a stringent length selection for 147 bp. Instead, a band from the gel that corresponds roughly to the length range $130 - 170$ bp was cut out [7]. Hence, the spectrum of the nucleosomal DNA lengths is at least as wide as the variation in the effective nucleosome size inferred by the HNG model (Fig. 2B) and the variation predicted by the SNG model (Fig. S15). The single-end sequencing yields one MNase cutting position from each nucleosomal DNA fragment, which reports one approximate position of one nucleosome endpoint. With DNA breathing at each end of the nucleosome and the stochastic nature of MNase cutting, the spacing between the cutting point and the nucleosome dyad is variable. However, this is equally true for every nucleosome at each end, and therefore should simply lead to a smearing of nucleosome positions. To study whether this affects our inferred effective nucleosome size (i.e. the parameter b within the HNG model), we used the data of Weiner *et al.* [8], which provides three levels of MNase digestion (typical, overdigested, and underdigested) for a single species (*S. cerevisiae*).

We performed our data analysis and fitting procedure identically on all three data sets and obtained $b = 143\,\mathrm{bp}$ for typical digestion, $b = 140\,\mathrm{bp}$ for overdigestion, and $b = 138\,\mathrm{bp}$ for underdigestion, see Fig. S19A-C and Table S2. The value for the typical digestion level compares well with the $b = 142\,\mathrm{bp}$ we obtained for the *S. cerevisiae* data set from the Tsankov *et al.* study. The variation of b with digestion level is weak and also does not display a trend. In contrast, the variation of b from species to species in Fig. 2B is significantly stronger and displays a clear trend. Furthermore, any unintentional variation in digestion level in the Tsankov *et al.* study would have been significantly less than the intentional variation in the Weiner *et al.* study, since the amount of MNase digestion for each species was determined by the same criterion in the Tsankov *et al.* study, thereby ensuring consistent digestion levels. Taken together, this indicates that our inference of b-values is robust and that the correlation observed in Fig. 2B is a real phenomenon and not an experimental or computational artifact. Note that the inference of the other parameter of Fig. 2B, the average repeat length, is also robust, since it is directly related to the period of the pattern.

As an additional independent test, we also tried to estimate an effective nucleosome size for each species in the Tsankov *et al.* data set directly from the primary data by performing

a cross-correlation analysis between the positions of the sequence reads on the Watson and the Crick strands. On average, the start points of sequence reads belonging to opposite ends of the same nucleosome should be shifted by a distance that may serve as a proxy for the effective nucleosome size. We therefore determined the distance Δx_{WC} that leads to the maximal cross-correlation and compared it to the effective nucleosome size b discussed above. As the plot in Fig. S19D shows, the two measures are indeed correlated (Pearson correlation coefficient 0.51 for all species, and 0.66 for all species except *K. lactis*). However, it is also clear from Fig. S19A that Δx_{WC} is not a quantitative proxy for b. A possible interpretation of this finding is that the MNase assay acts also as an experimental probe for nucleosome breathing, similar to the first experimental assay, which used cleavage by restriction enzymes as probes of the transient site-exposure within nucleosomal DNA [13].

[1] Chou T (2003) An exact theory of histone-DNA adsorption and wrapping. *Europhys Lett* 62:753–759.

[2] Teif VB, Ettig R, Rippe K (2010) A lattice model for transcription factor access to nucleosomal DNA. *Biophys J* 99:2597–2607.

[3] Mattis DC, ed. (1993) *The Many-Body Problem: An Encyclopedia of Exactly Solved Models in One Dimension* (World Scientific Publishing, Singapore).

[4] Baxter RJ (1982) *Exactly Solved Models in Statistical Mechanics* (Academic Press).

[5] Binney JJ, Dowrick NJ, Fisher AJ, Newman MEJ (1992) *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Oxford University Press).

[6] Chen Y (1987) Binding of n-mers to one-dimensional lattices with longer than close-contact interactions. *Biophys Chem* 27:59–65.

[7] Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ (2010) The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol* 8:e1000414.

[8] Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res* 20:90–100

[9] Kent WJ (2002) BLAT - the BLASt-like alignment tool. *Genome Res* 12:656–664.

[10] Möbius W, Gerland U (2010) Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLoS Comp Biol* 6:e1000891.

[11] Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.

[12] Celona B, Weiner A, Di Felice F, Mancuso FM, Cesarini E, et al. (2011) Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. *PLoS Biol* 9:e1001086.

[13] Polach KJ, Widom J (1995) Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J Mol Biol* 254:130–49.

Supplementary Tables

Model	**Parameter**	**Min. value**	**Max. value**	**Mesh spacing**
HNG (fixed b)	$b = 147\,\mathrm{bp}$	-	-	-
(for each species)	$1/\overline{\rho}$ [bp]	150	220	1
HNG (variable b)	b [bp]	120	$1/\overline{\rho} - 1$	1
(for each species)	$1/\overline{\rho}$ [bp]	150	220	1
SNG	ϵ [k_BT/bp]	0.1	0.17	0.0025
(for each species)	w [bp]	76	105	1
	μ [k_BT]	2	8	0.1
unified SNG	ϵ [k_BT/bp]	0.1	0.17	0.0025
(all species simultaneously,	w [bp]	76	105	1
except *K. lactis*)	$\mu_{S.\ cer.}$, ... [k_BT]	2	8	0.1
unified SNG	$\epsilon = 0.1525\ k_BT$/bp	-	-	-
(applied to *K. lactis*)	$w = 83\,\mathrm{bp}$	-	-	-
	$\mu_{K.\ lac.}$ [k_BT]	2	8	0.1

TABLE S1: Ranges and mesh sizes for optimizing parameters in the four different scenarios considered: HNG model with fixed and variable particle width b as well as the SNG model and the SNG model with simultaneous optimization (unified SNG model, *K. lactis* treated separately as indicated).

Species	**HNG model (fixed b)**		**HNG model (variable b)**		
	$1/\overline{\rho}$ [bp]	δ^2	b [bp]	$1/\overline{\rho}$ [bp]	δ^2
S. cerevisiae	171	$0.82 \cdot 10^{-6}$	142	170	$0.29 \cdot 10^{-6}$
S. paradoxus	169	$0.90 \cdot 10^{-6}$	143	168	$0.38 \cdot 10^{-6}$
S. mikatae	171	$1.22 \cdot 10^{-6}$	141	170	$0.42 \cdot 10^{-6}$
S. bayanus	171	$0.86 \cdot 10^{-6}$	144	171	$0.55 \cdot 10^{-6}$
C. glabrata	169	$0.95 \cdot 10^{-6}$	144	168	$0.69 \cdot 10^{-6}$
S. castellii	168	$0.43 \cdot 10^{-6}$	146	168	$0.37 \cdot 10^{-6}$
K. lactis	183	$2.50 \cdot 10^{-6}$	158	180	$0.57 \cdot 10^{-6}$
K. waltii	169	$4.10 \cdot 10^{-6}$	138	163	$0.63 \cdot 10^{-6}$
S. kluyverii	171	$0.48 \cdot 10^{-6}$	145	171	$0.37 \cdot 10^{-6}$
D. hansenii	171	$1.59 \cdot 10^{-6}$	140	168	$0.36 \cdot 10^{-6}$
C. albicans	182	$0.29 \cdot 10^{-6}$	149	182	$0.23 \cdot 10^{-6}$
Y. lipolytica	171	$0.36 \cdot 10^{-6}$	146	171	$0.33 \cdot 10^{-6}$
S. cerevisiae typical digestion	-	-	143	169	$0.41 \cdot 10^{-6}$
S. cerevisiae overdigestion	-	-	140	171	$0.31 \cdot 10^{-6}$
S. cerevisiae underdigestion	-	-	138	177	$0.18 \cdot 10^{-6}$

TABLE S2: Best-fit parameters for the HNG model (fit range starting at $x = 200$ bp, convolution over the range up to a quarter of the +1 peak height). These values are the basis for all plots in the main paper and this supplement.

Species	**SNG model**				**unified SNG model**			
	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2
S. cerevisiae	0.1550	81	166	$0.28{\cdot}10^{-6}$	0.1525	83	165	$0.32{\cdot}10^{-6}$
S. paradoxus	0.1625	81	165	$0.37{\cdot}10^{-6}$	0.1525	83	164	$0.37{\cdot}10^{-6}$
S. mikatae	0.1475	81	166	$0.41{\cdot}10^{-6}$	0.1525	83	164	$0.51{\cdot}10^{-6}$
S. bayanus	0.1625	81	167	$0.52{\cdot}10^{-6}$	0.1525	83	166	$0.53{\cdot}10^{-6}$
C. glabrata	0.1650	82	165	$0.68{\cdot}10^{-6}$	0.1525	83	165	$0.71{\cdot}10^{-6}$
S. castellii	0.1700	84	164	$0.33{\cdot}10^{-6}$	0.1525	83	165	$0.57{\cdot}10^{-6}$
K. lactis	0.1700	89	177	$0.41{\cdot}10^{-6}$	0.1525	83	179	$2.05{\cdot}10^{-6}$
K. waltii	0.1700	78	159	$0.60{\cdot}10^{-6}$	0.1525	83	158	$0.66{\cdot}10^{-6}$
S. kluyverii	0.1575	83	167	$0.35{\cdot}10^{-6}$	0.1525	83	167	$0.36{\cdot}10^{-6}$
D. hansenii	0.1475	81	163	$0.34{\cdot}10^{-6}$	0.1525	83	161	$0.41{\cdot}10^{-6}$
C. albicans	0.1275	88	175	$0.19{\cdot}10^{-6}$	0.1525	83	178	$0.20{\cdot}10^{-6}$
Y. lipolytica	0.1675	82	168	$0.30{\cdot}10^{-6}$	0.1525	83	168	$0.33{\cdot}10^{-6}$

TABLE S3: Best-fit parameters for the SNG model (fit range starting at $x = 200$ bp, convolution over the range up to a quarter of the +1 peak height). These values are the basis for all plots in the main paper and this supplement.

Species	**SNG model**				**unified SNG model**			
	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2
S. cerevisiae	0.1300	86	163	$0.34\cdot10^{-6}$	0.1400	85	163	$0.39\cdot10^{-6}$
S. paradoxus	0.1425	86	162	$0.41\cdot10^{-6}$	0.1400	85	163	$0.40\cdot10^{-6}$
S. mikatae	0.1275	86	163	$0.51\cdot10^{-6}$	0.1400	85	163	$0.63\cdot10^{-6}$
S. bayanus	0.1375	86	165	$0.63\cdot10^{-6}$	0.1400	85	166	$0.64\cdot10^{-6}$
C. glabrata	0.1475	85	164	$0.71\cdot10^{-6}$	0.1400	85	164	$0.72\cdot10^{-6}$
S. castellii	0.1625	84	165	$0.36\cdot10^{-6}$	0.1400	85	165	$0.63\cdot10^{-6}$
K. lactis	0.1600	90	177	$0.46\cdot10^{-6}$	0.1400	85	181	$1.68\cdot10^{-6}$
K. waltii	0.1500	81	158	$0.66\cdot10^{-6}$	0.1400	85	158	$0.81\cdot10^{-6}$
S. kluyverii	0.1400	87	166	$0.37\cdot10^{-6}$	0.1400	85	167	$0.38\cdot10^{-6}$
D. hansenii	0.1350	86	161	$0.40\cdot10^{-6}$	0.1400	85	161	$0.43\cdot10^{-6}$
C. albicans	0.1175	92	173	$0.23\cdot10^{-6}$	0.1400	85	179	$0.35\cdot10^{-6}$
Y. lipolytica	0.1425	86	167	$0.37\cdot10^{-6}$	0.1400	85	167	$0.38\cdot10^{-6}$

TABLE S4: As a test of the robustness of our results with respect to details of the fitting procedure: Best-fit parameters for the SNG model with fitting range starting at a distance of 100 bp from the +1 nucleosome, and convolution of the pattern over the shape of the +1 peak up to half of the peak height.

Species	**SNG model**				**unified SNG model**			
	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2
S. cerevisiae	0.1425	84	164	$0.30\cdot10^{-6}$	0.1525	83	165	$0.33\cdot10^{-6}$
S. paradoxus	0.1525	84	164	$0.39\cdot10^{-6}$	0.1525	83	164	$0.39\cdot10^{-6}$
S. mikatae	0.1375	84	164	$0.45\cdot10^{-6}$	0.1525	83	164	$0.56\cdot10^{-6}$
S. bayanus	0.1450	85	165	$0.60\cdot10^{-6}$	0.1525	83	166	$0.60\cdot10^{-6}$
C. glabrata	0.1625	83	164	$0.76\cdot10^{-6}$	0.1525	83	165	$0.79\cdot10^{-6}$
S. castellii	0.1700	84	164	$0.37\cdot10^{-6}$	0.1525	83	165	$0.62\cdot10^{-6}$
K. lactis	0.1700	89	177	$0.44\cdot10^{-6}$	0.1525	83	179	$2.21\cdot10^{-6}$
K. waltii	0.1600	80	159	$0.65\cdot10^{-6}$	0.1525	83	158	$0.68\cdot10^{-6}$
S. kluyverii	0.1500	86	165	$0.38\cdot10^{-6}$	0.1525	83	167	$0.39\cdot10^{-6}$
D. hansenii	0.1425	85	161	$0.40\cdot10^{-6}$	0.1525	83	162	$0.46\cdot10^{-6}$
C. albicans	0.1250	91	173	$0.24\cdot10^{-6}$	0.1525	83	179	$0.31\cdot10^{-6}$
Y. lipolytica	0.1550	84	168	$0.33\cdot10^{-6}$	0.1525	83	168	$0.36\cdot10^{-6}$

TABLE S5: As a test of the robustness of our results with respect to details of the fitting procedure: Best-fit parameters for the SNG model with fitting range starting at a distance of 100 bp from the +1 nucleosome, and convolution of the pattern over the shape of the +1 peak up to a quarter of the peak height.

Species	**SNG model**				**unified SNG model**			
	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2	ϵ $[k_BT/\text{bp}]$	w [bp]	$1/\overline{\rho}$ [bp]	δ^2
S. cerevisiae	0.1425	82	166	$0.27{\cdot}10^{-6}$	0.1475	83	165	$0.33{\cdot}10^{-6}$
S. paradoxus	0.1500	82	165	$0.35{\cdot}10^{-6}$	0.1475	83	164	$0.35{\cdot}10^{-6}$
S. mikatae	0.1375	82	166	$0.41{\cdot}10^{-6}$	0.1475	83	164	$0.53{\cdot}10^{-6}$
S. bayanus	0.1500	82	167	$0.52{\cdot}10^{-6}$	0.1475	83	166	$0.53{\cdot}10^{-6}$
C. glabrata	0.1575	82	166	$0.63{\cdot}10^{-6}$	0.1475	83	165	$0.65{\cdot}10^{-6}$
S. castellii	0.1675	83	165	$0.29{\cdot}10^{-6}$	0.1475	83	165	$0.54{\cdot}10^{-6}$
K. lactis	0.1625	90	177	$0.40{\cdot}10^{-6}$	0.1475	83	178	$2.23{\cdot}10^{-6}$
K. waltii	0.1575	79	159	$0.58{\cdot}10^{-6}$	0.1475	83	157	$0.62{\cdot}10^{-6}$
S. kluyverii	0.1450	84	167	$0.34{\cdot}10^{-6}$	0.1475	83	168	$0.34{\cdot}10^{-6}$
D. hansenii	0.1425	81	164	$0.34{\cdot}10^{-6}$	0.1475	83	162	$0.43{\cdot}10^{-6}$
C. albicans	0.1225	88	176	$0.19{\cdot}10^{-6}$	0.1475	83	178	$0.22{\cdot}10^{-6}$
Y. lipolytica	0.1625	82	169	$0.30{\cdot}10^{-6}$	0.1475	83	168	$0.33{\cdot}10^{-6}$

TABLE S6: As a test of the robustness of our results with respect to details of the fitting procedure: Best-fit parameters for the SNG model with fitting range starting at a distance of 200 bp from the +1 nucleosome, and convolution of the pattern over the shape of the +1 peak up to half of the peak height.

Supplementary Figures

S. cerevisiae

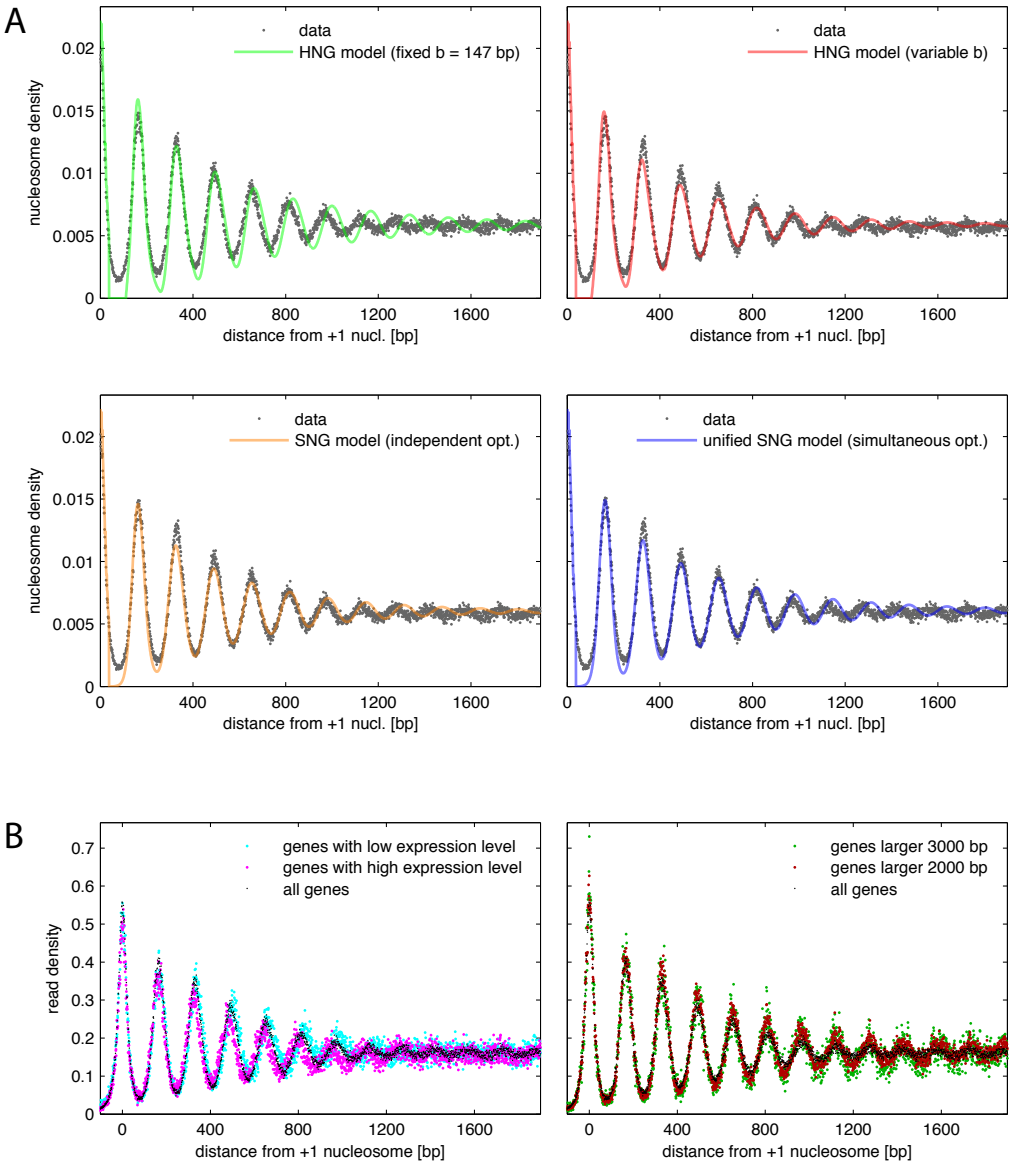

FIG. S1: *S. cerevisiae*: (A) Nucleosome density aligned to +1 nucleosome and averaged over many genes (gray dots, normalized using estimated sequencing depth) and HNG model with fixed $b = 147\,\mathrm{bp}$ (green line), HNG model with variable b (red line), SNG model (orange line), and unified SNG model (blue line) with parameters optimized (see Tables S2 and S3). (B) Read density when averaging over all genes (black) and when averaging over those 20 % with lowest (cyan) / highest (magenta) expression level or those genes which are longer than 3000 bp (green) / 2000 bp (red).

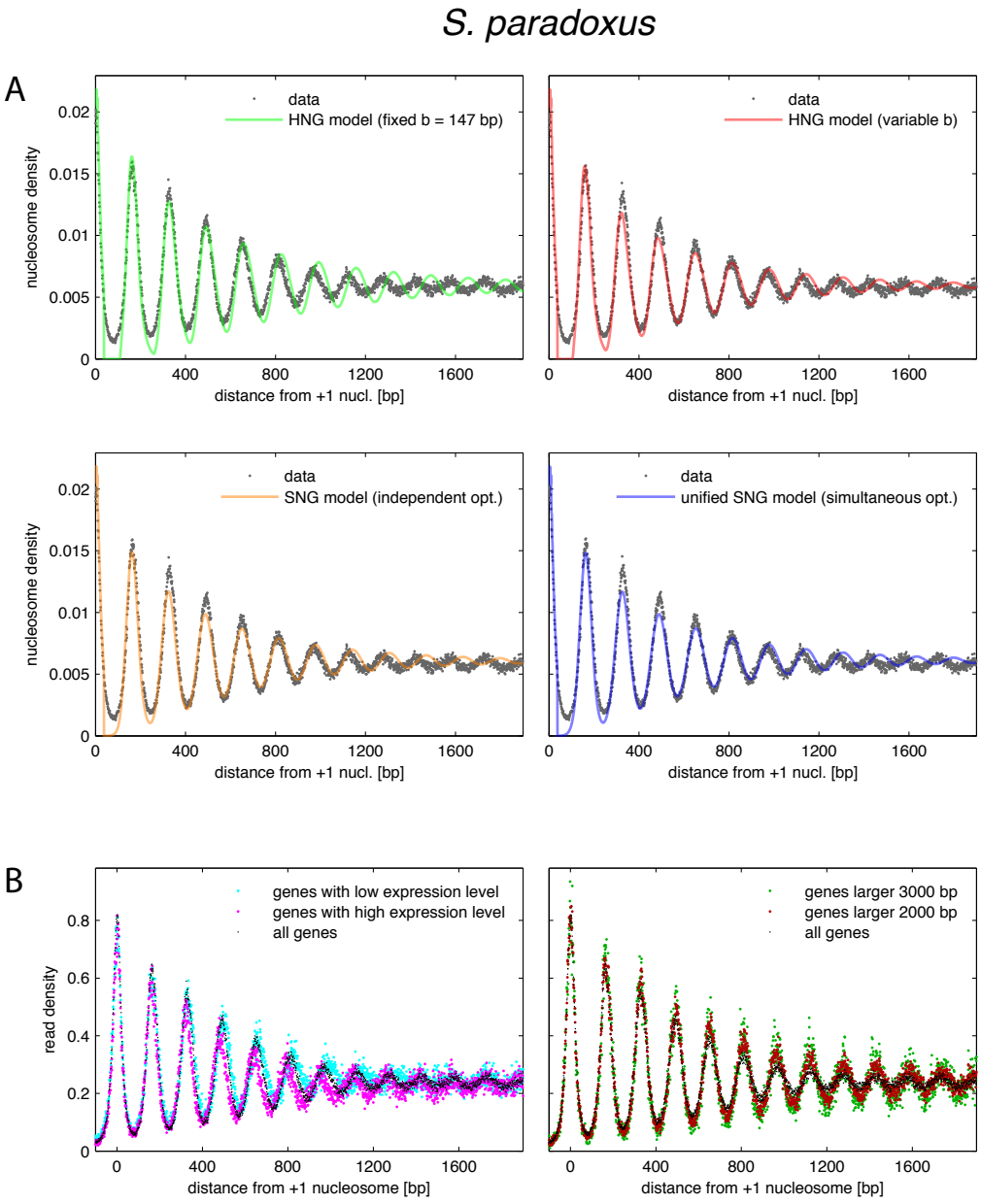

FIG. S2: Like Figure S1, but for *S. paradoxus*.

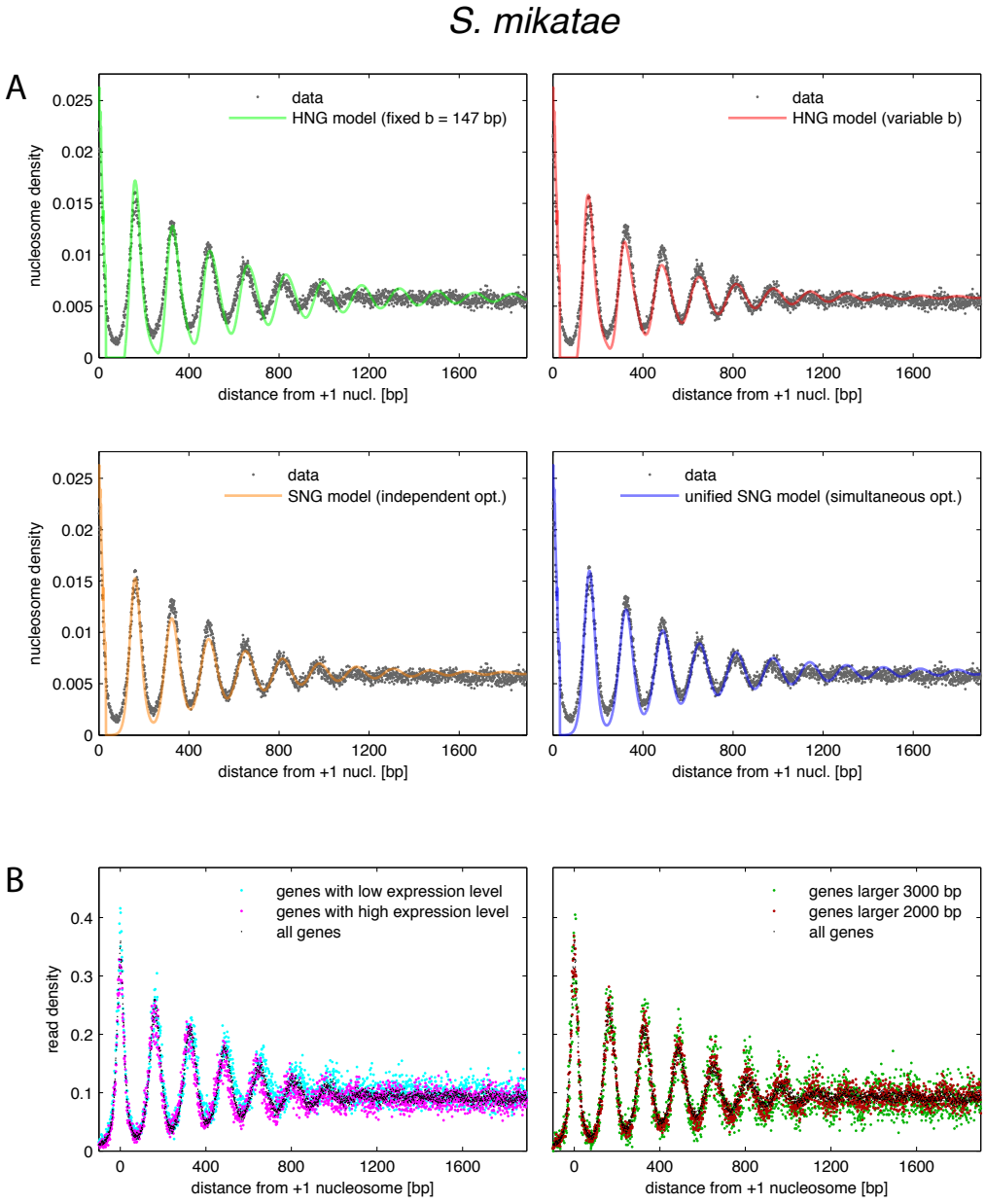

FIG. S3: Like Figure S1, but for *S. mikatae*.

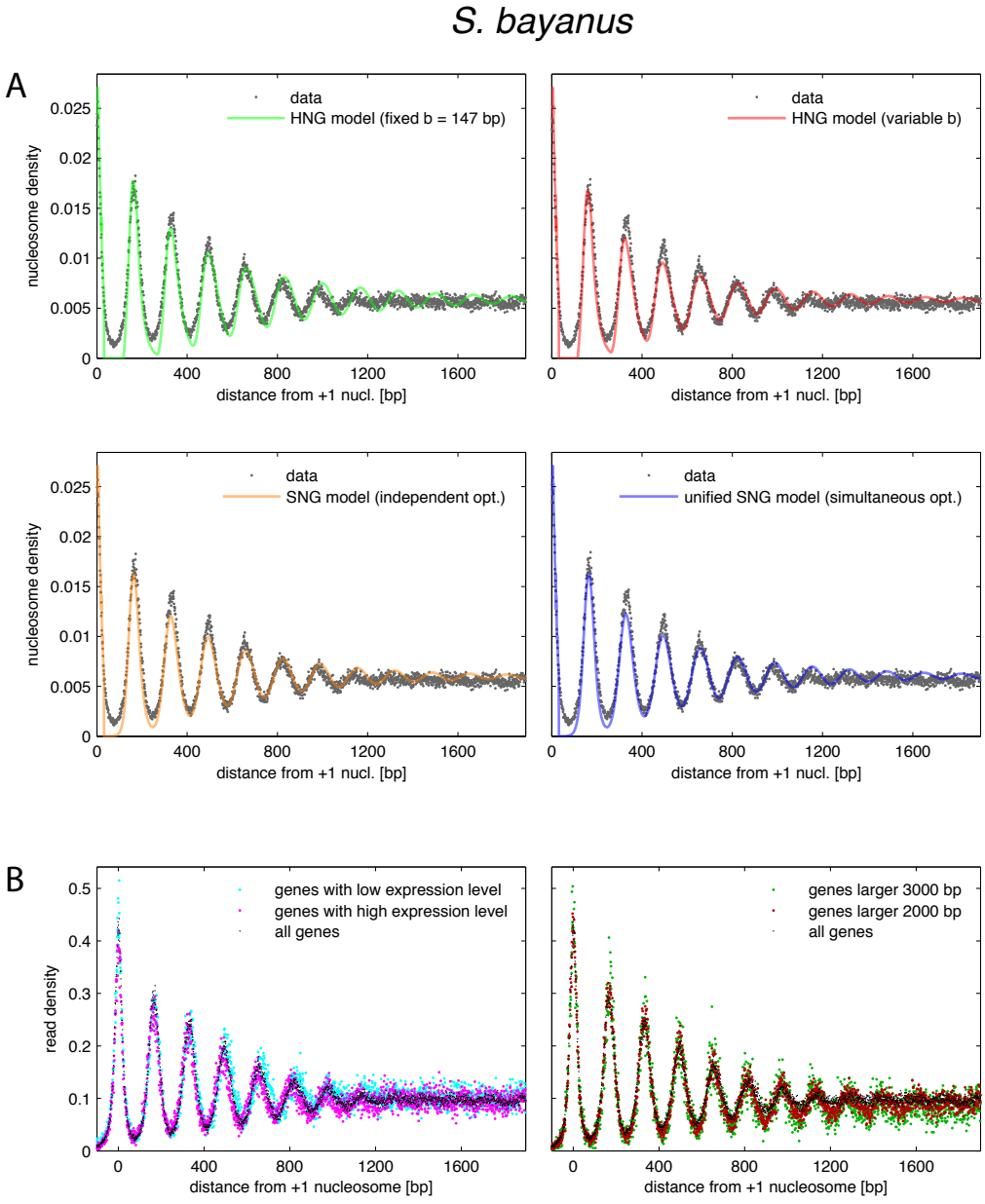

FIG. S4: Like Figure S1, but for *S. bayanus*.

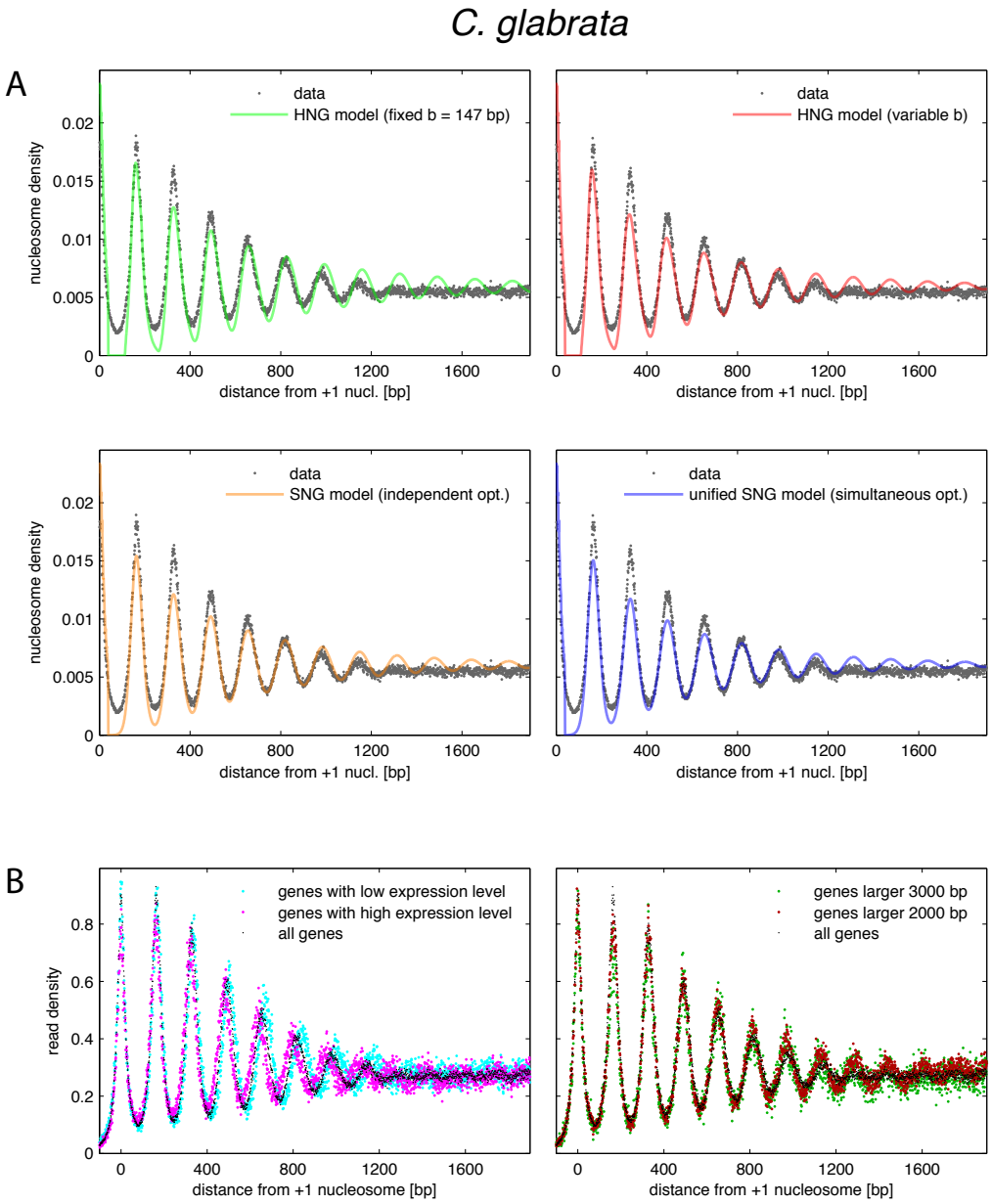

FIG. S5: Like Figure S1, but for *C. glabrata*.

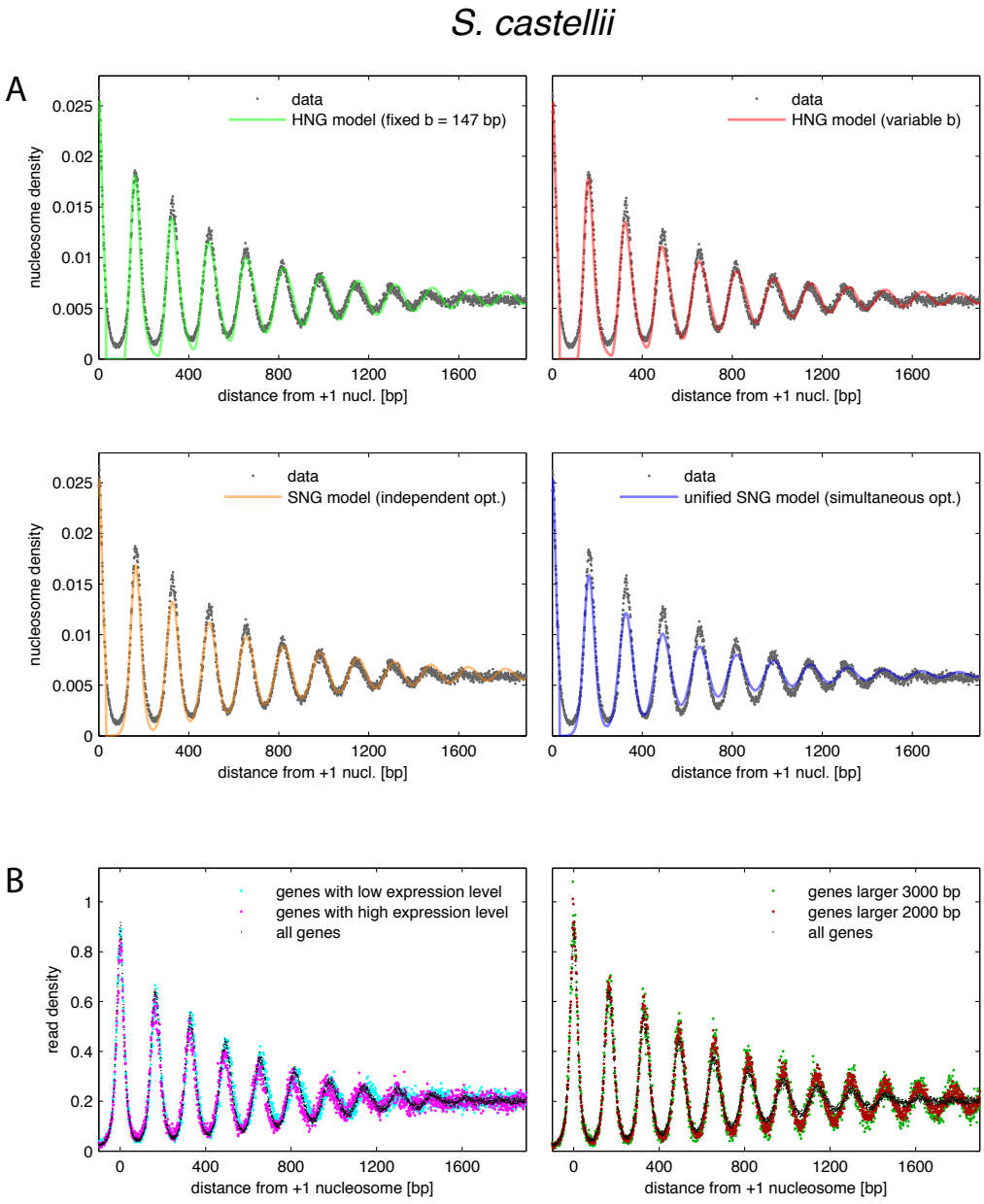

FIG. S6: Like Figure S1, but for *S. castellii*.

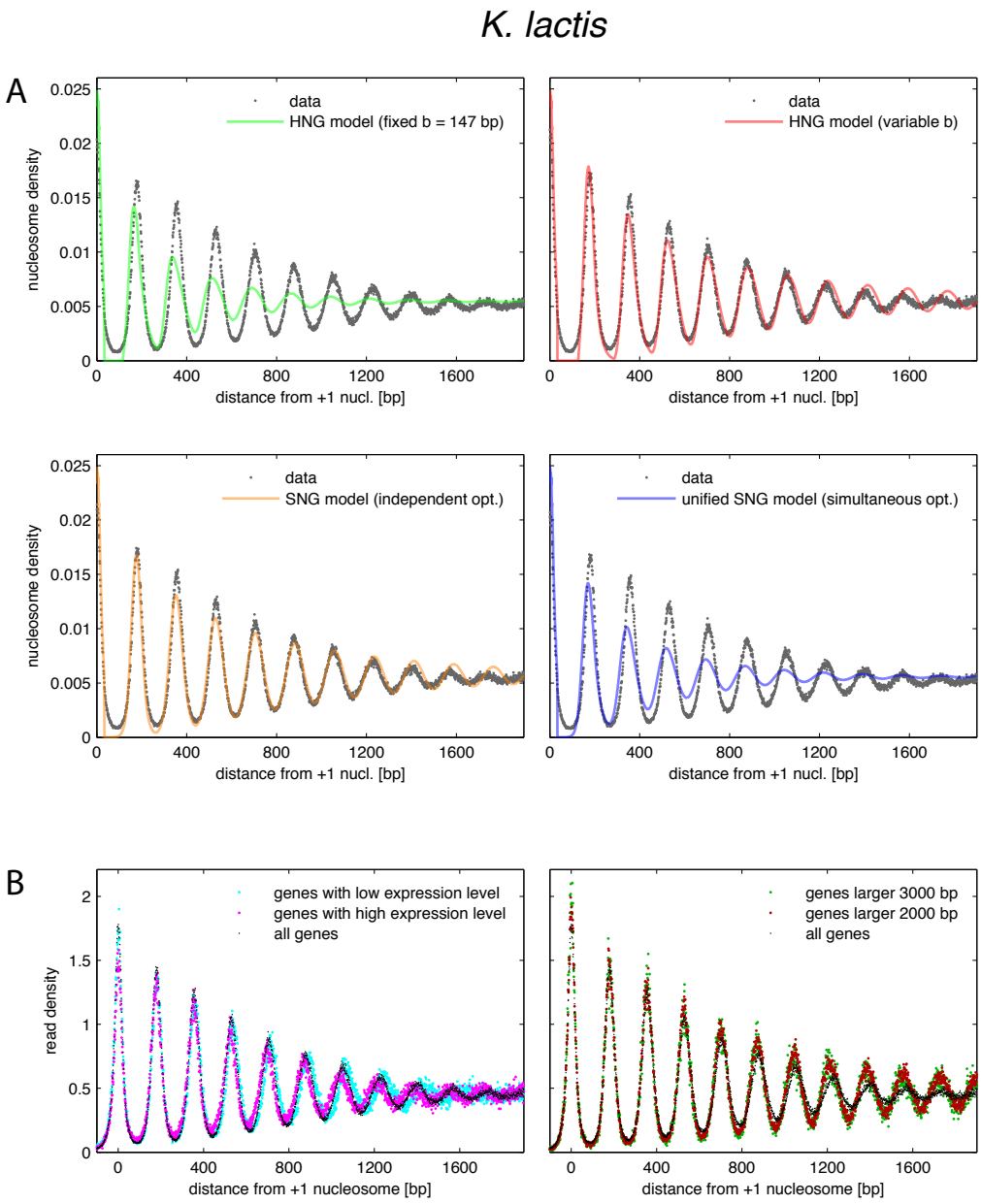

FIG. S7: Like Figure S1, but for *K. lactis*, except lower right of panel (A): Here, the SNG model is applied with parameters obtained from the fit to all species simultaneously except *K. lactis*, see Table S1 for more details. The mismatch illustrates the fact that the unified model cannot explain the data from *K. lactis*.

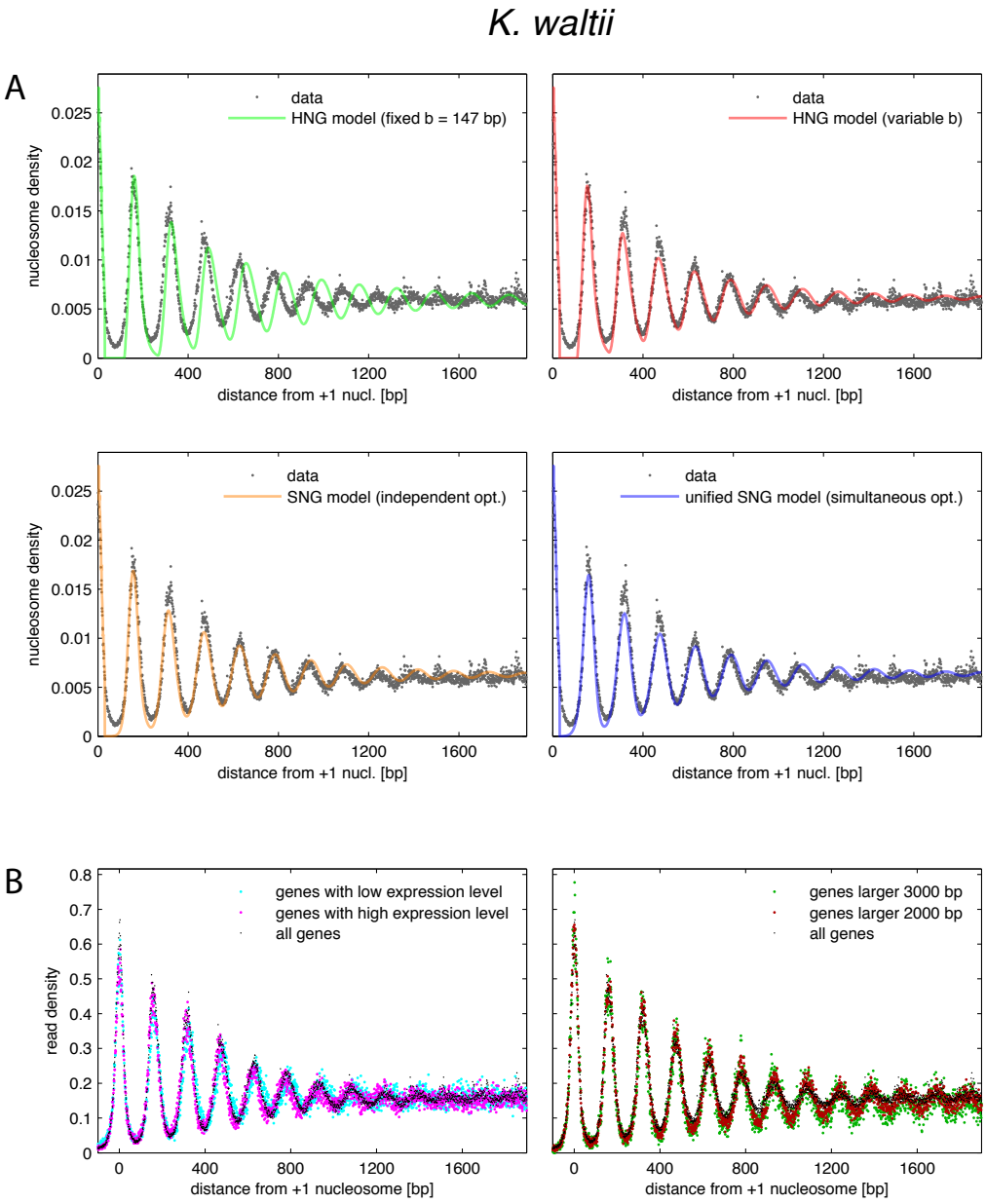

FIG. S8: Like Figure S1, but for *K. waltii*.

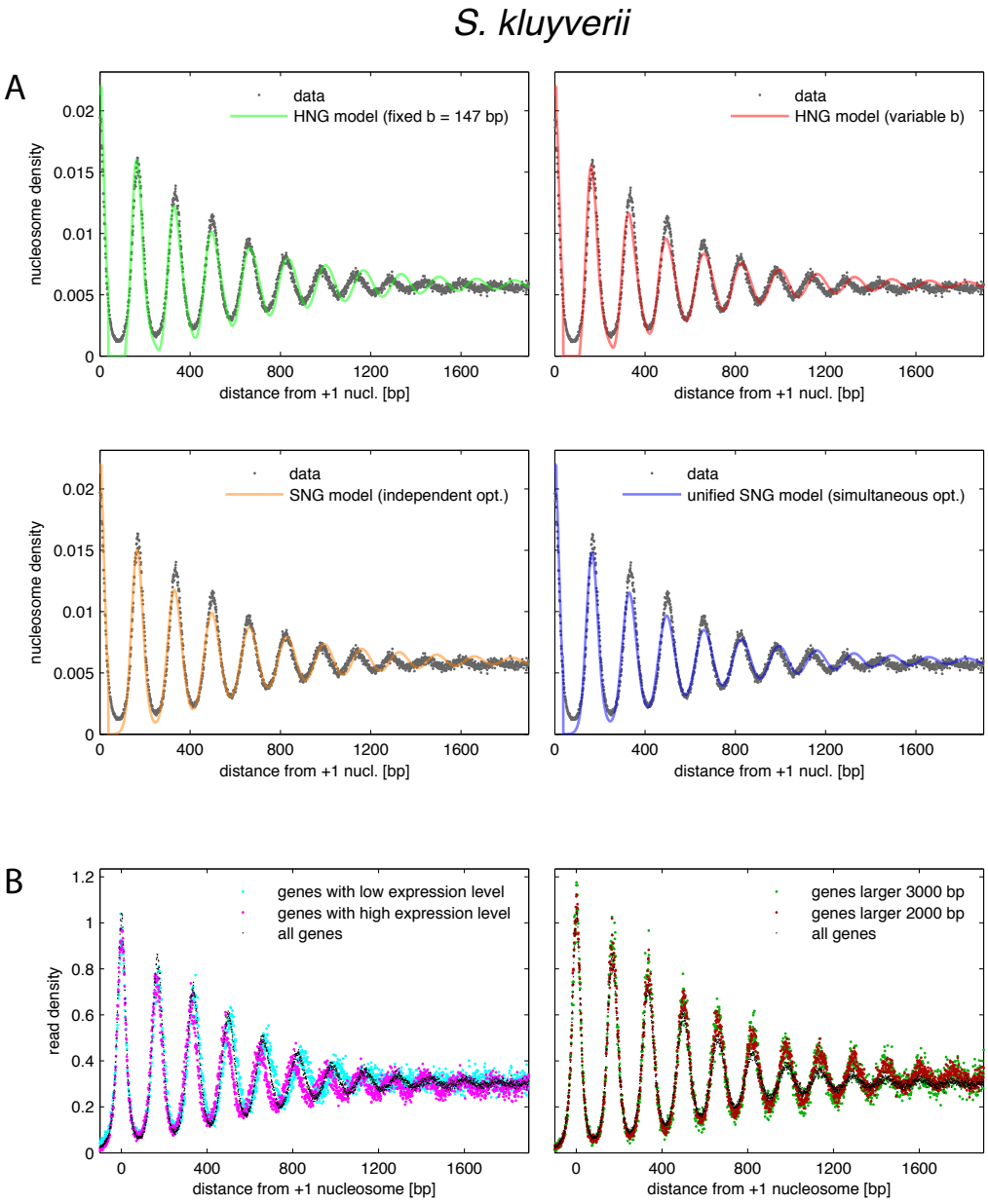

FIG. S9: Like Figure S1, but for *S. kluyverii*.

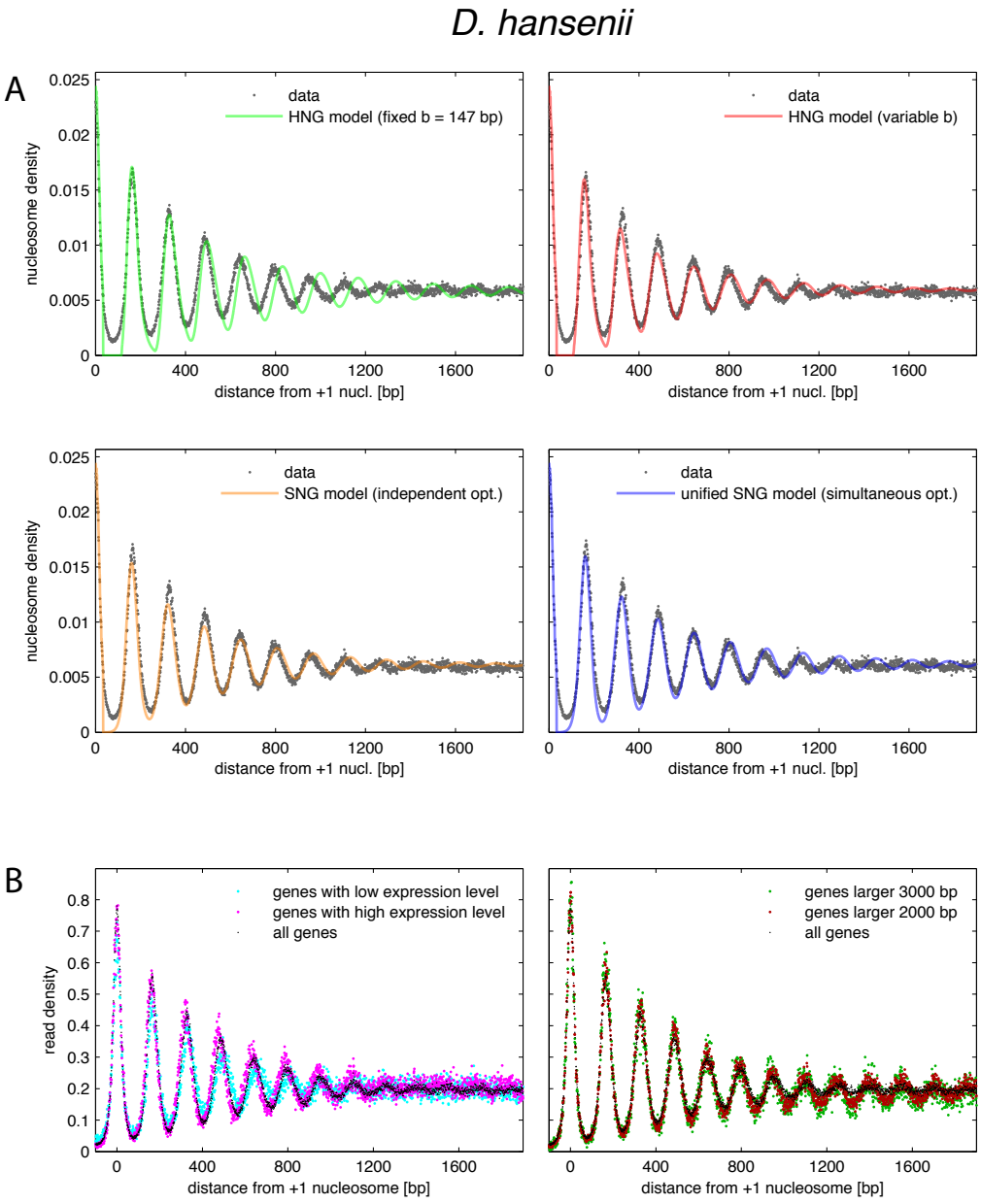

FIG. S10: Like Figure S1, but for *D. hansenii*.

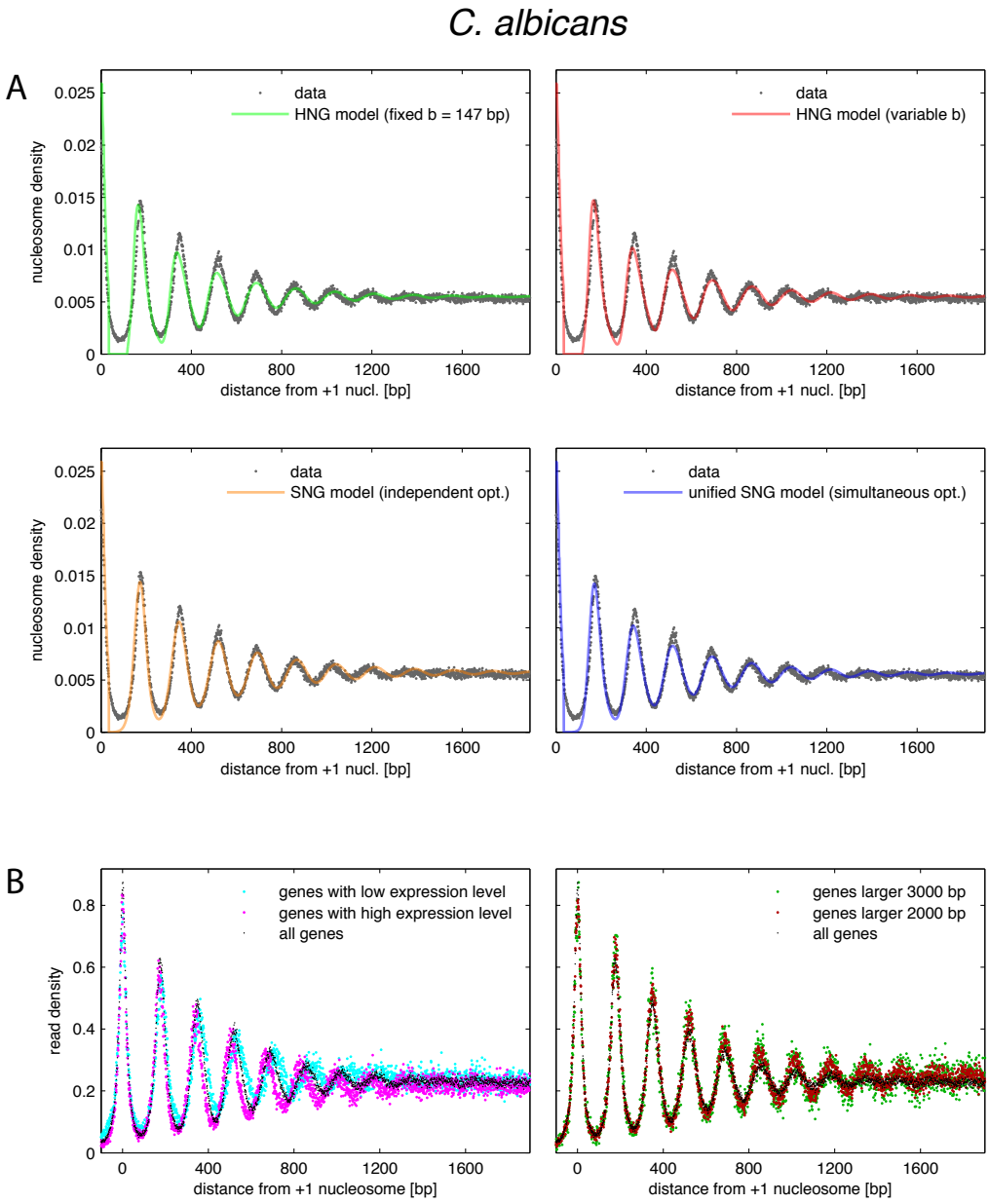

FIG. S11: Like Figure S1, but for *C. albicans*.

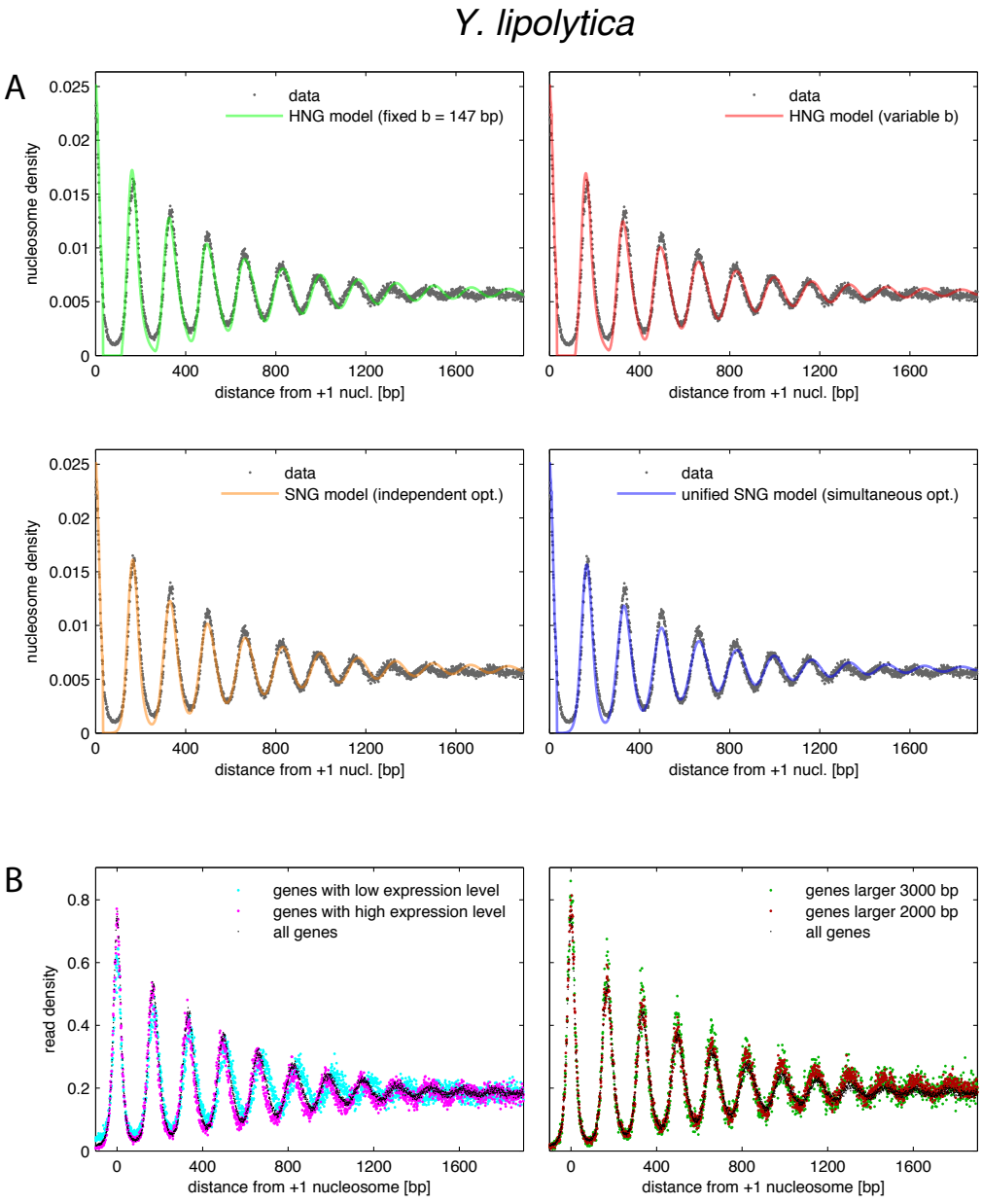

FIG. S12: Like Figure S1, but for *Y. lipolytica*.

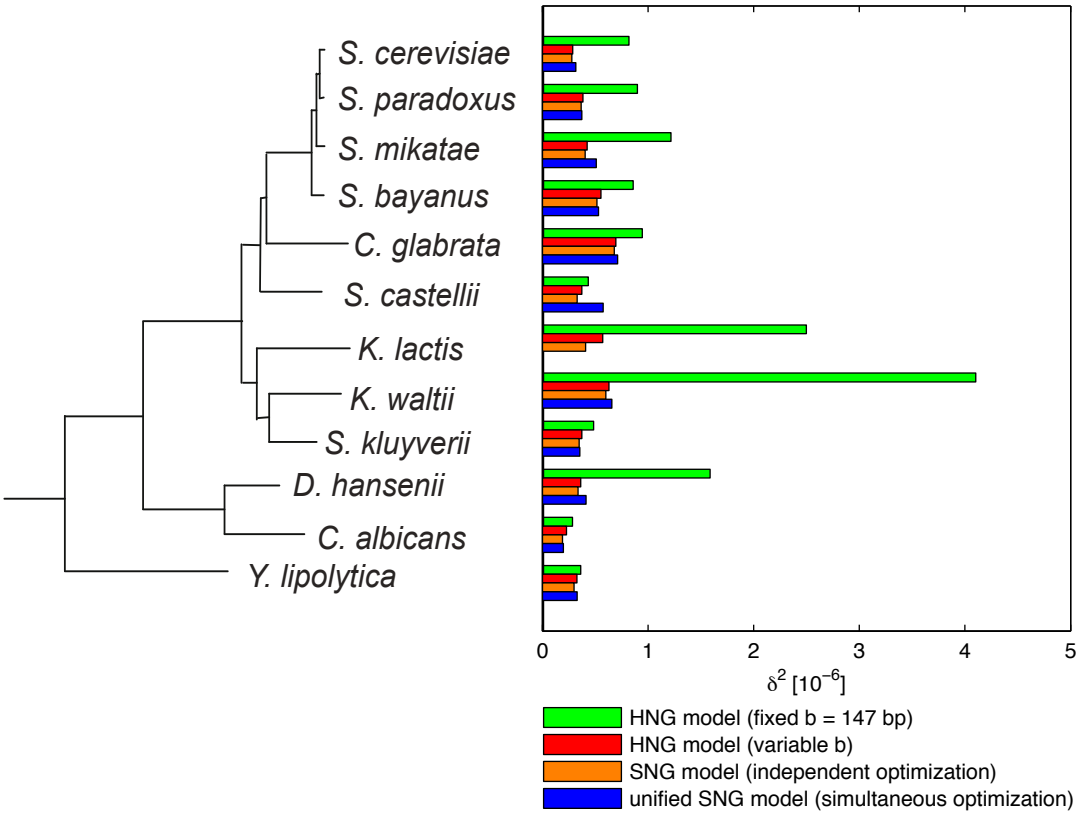

FIG. S13: Comparison of how well the different models considered describe the data. For each model and each species the mean squared deviation per data point, δ^2, is displayed. *K. lactis* is not included in the unified SNG model, as explained in the main text.

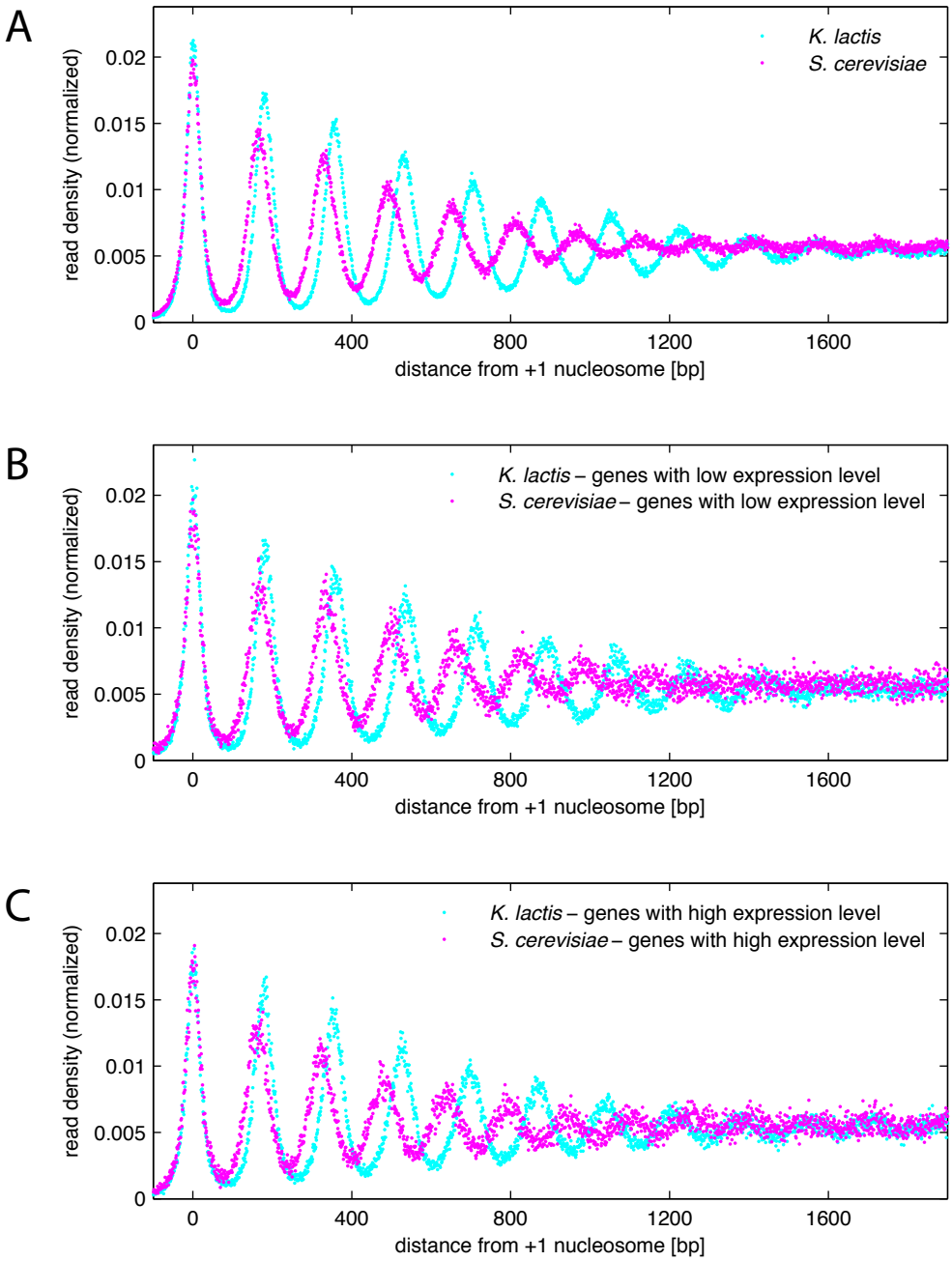

FIG. S14: Read density averaged over (A) all genes, (B) those 20% of genes with lowest and (C) those 20% with highest expression level for *K. lactis* (cyan) and *S. cerevisiae* (magenta). Normalization has been introduced to account for different coverage among both species.

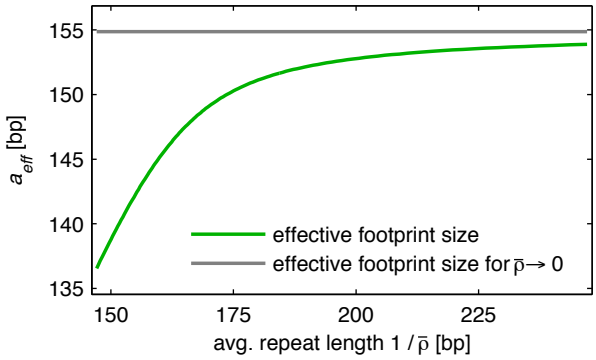

FIG. S15: Pressure-induced unwrapping effect in the SNG model. Effective footprint size a_{eff} as a function of average repeat length $1/\overline{\rho}$ for our consensus parameter values of $\epsilon = 0.1525\, k_B T/\text{bp}$, $w = 83\,\text{bp}$ (green line). Also indicated is the limiting value for large $1/\overline{\rho}$ (gray line).

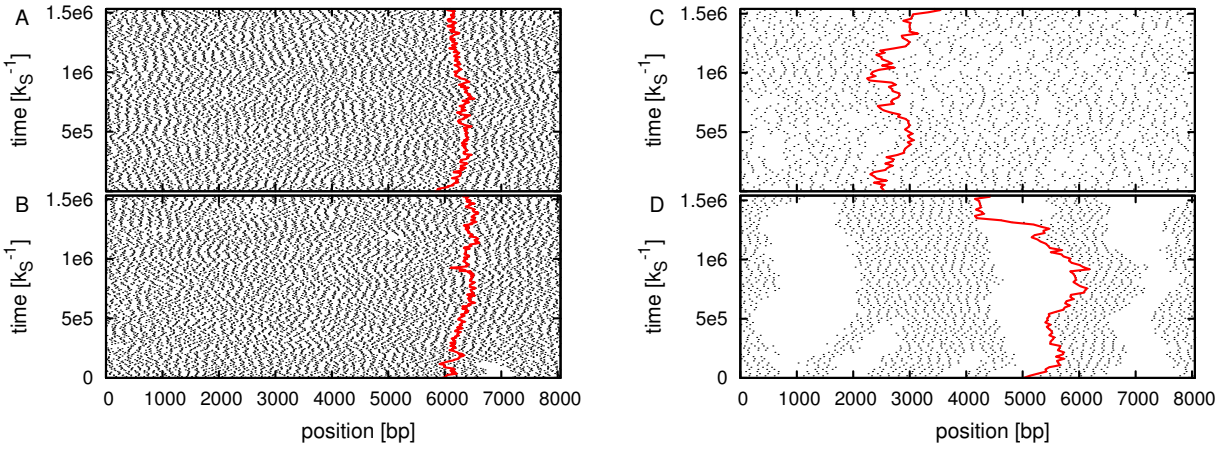

FIG. S16: Active SNG model. Kymograph of nucleosome positions (black dots) in a periodic system with fixed number of nucleosomes corresponding to an inverse density of (A),(B) $1/\overline{\rho} = 175$ bp and (C),(D) $1/\overline{\rho} = 300$ bp, and remodeler activity (A),(C) $k_{rm} = 0$ and (B),(D) $k_{rm} = 0.08$, respectively. Remodeling evidently leads to clustering at low-density, while little change is observed at high-density. $k_s = 1$ serves as the basic time unit. In each case, the trajectory of a single, randomly selected particle is traced in red.

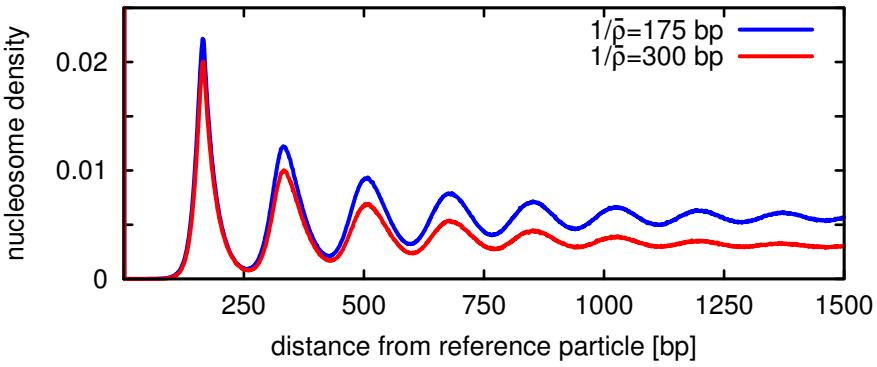

FIG. S17: Active SNG model. Nucleosome density close to a reference particle in a periodic system with fixed number of nucleosomes corresponding to the inverse density $1/\overline{\rho}$ as indicated and a remodeler rate of $k_{rm} = 0.08$. Note the qualitative similarity with Figure 7 of Ref. [12].

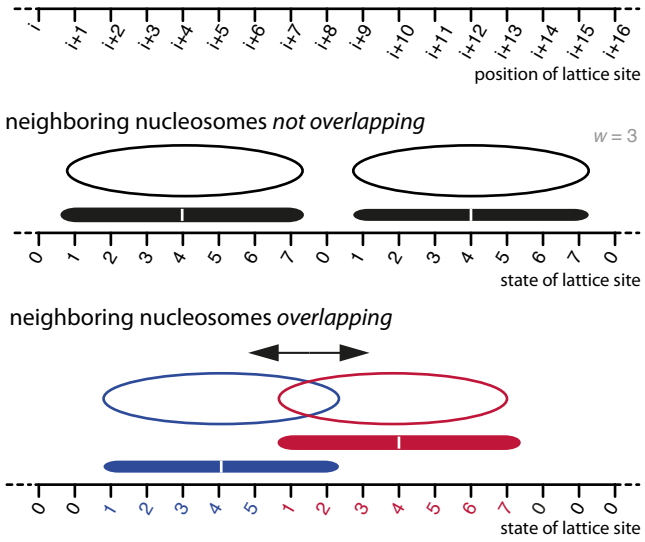

FIG. S18: Illustration of the state definitions for each site for nucleosome pairs either overlapping or non-overlapping.

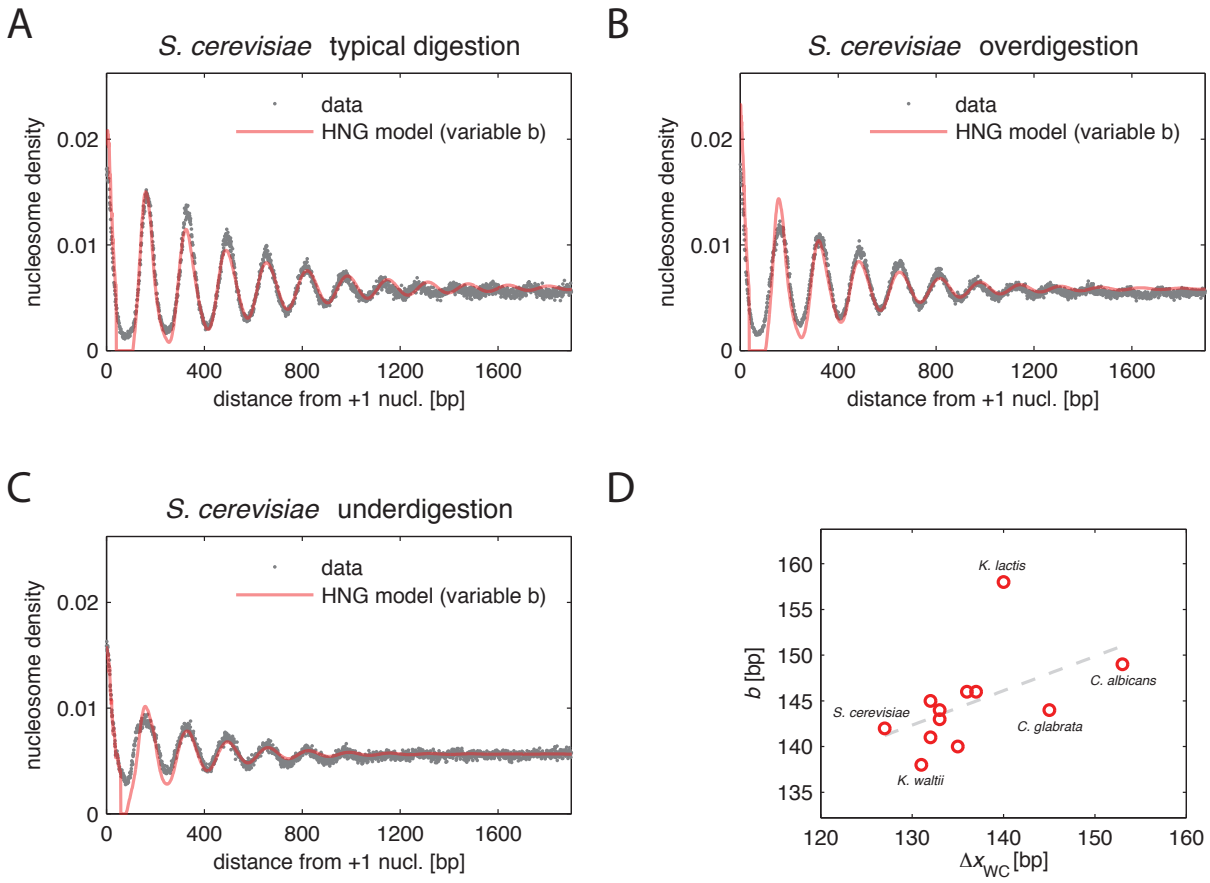

FIG. S19: Additional analysis to exclude MNase digestion artefacts in the estimation of effective nucleosome widths b based on the HNG model. See Section VI for a detailed description.