(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 February 2018 (08.02.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/024749 A1

(51) International Patent Classification:
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/EP2017/069480

(22) International Filing Date:
01 August 2017 (01.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16382379.2 01 August 2016 (01.08.2016) EP

(71) Applicants: **CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS** [ES/ES]; Calle Serrano 117, 28006 Madrid (ES). **UNIVERSIDAD DE SALAMANCA** [ES/ES]; Patio de Escuelas, 1, 37008 Salamanca (ES).

(72) Inventors: **ANTEQUERA MÁRQUEZ, Francisco**; Consejo Superior de Investigaciones Científicas, Instituto de Biología Funcional y Genómica, Calle Zacarías González 2, 37007 Salamanca (ES). **MIGUEL QUINTALES, Luis Antonio**; Universidad de Salamanca, Instituto de Biología Funcional y Genómica, Calle Zacarías González 2, 37007 Salamanca (ES).

(74) Agent: **HOFFMANN EITLE S.L.U.**; Paseo de la Castellana 140. Planta 3ª., Edificio Lima, 28046 Madrid (ES).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: A METHOD FOR TAILORING A DNA SEQUENCE TO OBTAIN SPECIES-SPECIFIC NUCLEOSOME POSITIONING

(57) Abstract: The present invention relates to a method for obtaining the DNA sequence pattern determining the nucleosome positioning characteristic of a particular eukaryotic species. It further relates to a method for tailoring coding and/or non-coding DNA sequences to obtain the nucleosome positioning characteristic of a given eukaryotic species. In addition, it also relates to a method of codon optimization based on the modification of nucleosome positioning of a coding and/or non-coding DNA sequence for improving its expression in a host eukaryotic species.

WO 2018/024749 A1

A METHOD FOR TAILORING A DNA SEQUENCE TO OBTAIN SPECIES-SPECIFIC NUCLEOSOME POSITIONING

FIELD OF INVENTION

The present invention can be included in the field of genetic engineering and genomics. In particular, it relates to a method for obtaining the DNA sequence pattern determining the nucleosome positioning characteristic of a particular eukaryotic species. It further relates to a method for tailoring coding and/or non-coding DNA sequences to obtain the nucleosome positioning characteristic of a given eukaryotic species. In addition, it also relates to a method of codon optimization based on the modification of nucleosome positioning of a coding and/or non-coding DNA sequence for improving its expression in a host eukaryotic species.

BACKGROUND OF THE INVENTION

Eukaryotic nuclear DNA is packaged into nucleosomes. The nucleosome core particle consists of ~147 base pairs (bp) of DNA wrapped ~1.7 times in a left-handed superhelical turn around an octamer of histone proteins—two each of H2A, H2B, H3, and H4 (Hughes and Rando 2014; Lieleg et al. 2015). This nucleosome core structure is well conserved and impedes DNA access for most factors.

Most nucleosomes occupy well-defined positions along a eukaryotic species genome. These have been described in yeast to remain constant under many different physiological conditions (Yuan et al. 2005; Lee et al. 2007; Zhang et al. 2011a; Soriano et al. 2013). This precise positioning has been reported to be essential to modulate the access of proteins to specific sites in the chromosomes to regulate genomic processes (Hughes and Rando 2014), such as transcription (Bai et al. 2010; Koster et al. 2015), replication initiation (Lipford and Bell, 2001; Eaton et al. 2010; Berbenetz et al. 2010; Soriano et al. 2014) and recombination (Pan et al. 2011; de Castro et al. 2012).

Nucleosomal patterns have been reported to result from the combined contribution of chromatin remodelers, DNA-binding proteins, and the differential affinity of nucleosomes for different DNA sequences. Chromatin remodelers are multiprotein complexes that use ATP hydrolysis to facilitate the sliding, eviction or histone exchange of nucleosomes (Clapier and Cairns, 2009). Remodelers show different specificity and directionality in their mode of action (Stockdale et al. 2006; Yen et al. 2012) and the removal of some of them like Hrp3 in *Schizosaccharomyces pombe* (Pointner et al. 2012; Shim et al. 2012; Hennig et al. 2012) or Isw1 and Chd1 in *Saccharomyces cerevisiae* (Gkikopoulos et al. 2011) result in gross genome-wide alteration of their nucleosomal patterns. Transcription factors contribute to chromatin organization through

the recruitment of remodelers to promoters (Cosma et al. 1999; Yudkovsky et al. 1999; Korber et al. 2004) and also through their ability to compete with nucleosomes for their binding sites (Badis et al. 2008; Hartley and Madhani, 2009; Tsankov et al. 2011; Soriano et al. 2013; Koster et al. 2015). Transcriptional regulatory complexes bound at promoters could also act as physical barriers from which regular nucleosomal arrays are generated, as proposed by the statistical positioning model (Kornberg and Stryer, 1988; Mavrich et al. 2008) through active ATP-dependent mechanisms (Zhang et al. 2011b). The third element contributing to nucleosome positioning is the DNA sequence. The DNA molecule is strongly bent along its axis and adjacent nucleotides are also under strong lateral displacement to accommodate 147 bp of double stranded DNA in 1.7 turns around the histone octamer. Sequence motifs vary in their resistance to deformation and, therefore, different DNA molecules offer a different resistance to bending (Drew and Travers, 1985; Thastrom et al. 1999). Sequence analyses of aggregated nucleosomal profiles have revealed that some AT-rich dinucleotides are preferentially positioned in the minor groove of DNA facing the histone core while GC-rich dinucleotides face outwards (Satchwell et al. 1986; Ioshikhes et al. 1996; Lowary and Widom 1998; Segal et al. 2006; Albert et al. 2007). This alternating organization favours the bending of the DNA molecule and the electrostatic interaction between arginine residues and AT-rich sequences in the minor groove of DNA around the histone core (Rohs et al. 2009).

It is known that when exogenous DNA sequences are introduced into a host genome, these adopt a nucleosomal organization different from that of the endogenous genes (Zhang Z et al. 2011b). This can be a problem for the efficient expression of heterologous genes since the precise organization of nucleosomes on DNA is known to have an important role in the correct expression and stability of the genes.

There is a long-running debate over the extent to which DNA sequence determines nucleosome positioning *in vivo*. The search for sequence determinants of nucleosome positioning has led to the identification of some synthetic sequences with great affinity to form nucleosomes *in vitro*, (Lowary and Widom 1998) among which the 601 sequence has been extensively used in many structural studies (Olson and Zhurkin 2011; Ngo et al. 2015 and references therein). Natural sequences like the 5S RNA also have been described to have a strong positioning potential *in vitro* and *in vivo* (Simpson and Stafford, 1983; Pennings et al. 1991).

Despite these preferences, the extent to which the DNA sequence contributes to nucleosome positioning in the genomic context remains unclear. In some cases, the nucleosomal pattern *in vitro* coincides with that of native chromatin in discrete genomic regions (Shen and Clark, 2001; Allan et al. 2013; Beh et al. 2015). *In vitro* approaches however, do not always mimic the

situation *in vivo*. Indeed, a different positioning potential, *in vivo* and *in vitro*, has been reported for the same DNA sequence. For instance, unlike the situation *in vitro*, the 601 and 603 artificial sequences do not preferentially form nucleosomes when integrated into the genome of *S. cerevisiae* (Perales et al. 2011; Gaykalova et al. 2011). Other studies have shown that the DNA sequence is unable to recapitulate the *in vivo* positioning pattern at the genomic scale in chromatin assembly assays *in vitro* (Zhang et al. 2009).

Analysis of the substitution rate of mononucleosomal DNA in related species has previously suggested that nucleosomal positioning relative to the DNA sequence has remained stable over evolutionary timescales (Washieldt et al. 2008). This long-term association between histones and DNA makes possible that nucleosomal signatures could have emerged as a consequence of a different rate of mutation or biased repair along mononucleosomal and linker DNA, due to small differences in the structure of histone octamers or in the bias or accessibility of repair proteins (Washietl et al. 2008; Sasaki et al. 2009). Such a mutational scenario is compatible with the finding that sequence variation along mononucleosomal DNA is under positive selection in humans (Prendergast and Semple, 2011) and in *S. cerevisiae* (Warnecke et al. 2008), and suggests that nucleosomal signatures could have a positive selective value for their contribution to nucleosome positioning.

The authors of the invention previously described that the four nucleotides follow well-defined and asymmetrical patterns along mononucleosomal DNA (Quintales et al. 2015a). More specifically, the authors generated mono-, di-, trinucleotide and amino acid profiles by aligning mononucleosomal sequences 150 bp long associated with well-positioned nucleosomes to the nucleosome midposition (dyad) and calculating the frequencies of mononucleotides (figure 1 and electronic supplementary material, figure S2, Quintales et al. 2015a) for each position. Similarly, the frequencies of di- (electronic supplementary material, figure S4, Quintales et al. 2015a) and trinucleotides (figure 2 and electronic supplementary material, figures S5 and S6, Quintales et al. 2015a) and those of the sum of trinucleotides corresponding to codons for each amino acid (figures 2 and 3, and electronic supplementary material, figures S6 and S7, Quintales et al. 2015a) were also calculated for each position and normalized to the corresponding genome averages.

This nucleosomal signature was described by the authors to be present in transcribed and non-transcribed regions and was found to be species-specific, varying widely even among species of the same genus (Quintales et al. 2015a). Moreover, the authors disclosed that in the case of coding regions, the relative distribution of the 20 amino acids along proteins was correctly

predicted based on the position of the corresponding codons along the mononucleosomal DNA (Quintales et al. 2015a).

SUMMARY OF THE INVENTION

Accordingly, as summarized by Pugh (Pugh, B. et al. 2010) underlying DNA sequences have been reported not to be widespread determinants of nucleosome positioning *in vivo*, although they are major determinants at some positions. An important unresolved question is how nucleosomes become uniformly spaced and precisely positioned *in vivo*.

Despite some DNA sequences having been described to contribute to nucleosome positioning, on the basis of any of these sequences it has not been reported so far any method enabling the targeted positioning of nucleosomes *in vivo* in a DNA sequence, let alone to obtain a species-specific nucleosome positioning.

For the first time, the authors of the invention have shown that the species-specific characteristic DNA pattern disclosed in Quintales et al. 2015a, i.e., the species-specific nucleosomal signature, contains information capable of targeting nucleosomes to specific positions in coding and non-coding DNA sequences in order to adopt a species-specific nucleosome positioning pattern. Indeed, it has been found by the inventors that said positioning information is only correctly interpreted by the species from which a given DNA sequence is derived (e.g., Example 5).

To extract the sequence information contained in the species-specific signature, the inventors generated a matrix referred as position-specific weight matrix (PSWM) which is characteristic of each eukaryotic species wherein a score is determined for each position (i) in the mononucleosomal DNA sequence based on the frequency of a k-mer (e.g. a dinucleotide) at said position in the mononucleosomal DNA sequence with respect to the frequency of said oligonucleotide in the genome of said species.

The replacement of each oligonucleotide position in a heterologous DNA sequence for the oligonucleotide having the highest score in the species-specific PSWM for the corresponding position has resulted in nucleosome positioning profiles substantially identical to those of native sequences. This is shown in Example 4 for non-coding regions.

In coding regions, taking advantage of the degeneracy of the genetic code, codons along an heterologous mononucleosomal DNA sequence were replaced by synonymous codons with the highest possible score at each position of the PSWM generated for a particular yeast species

(i.e., *S. pombe*). Replacing the native codons by the synonymous codons generated regular nucleosomal arrays with a profile mostly indistinguishable from that of the endogenous coding sequence both for eukaryotic (Example 6) and prokaryotic (Example 7) sequences.

The obtained results further support the use of the information within nucleosomal signatures in a method for tailoring or customizing an heterologous DNA sequence for obtaining the nucleosomal organization characteristic of a given eukaryotic species and more specifically, for coding sequences a method of codon optimization, based on the modification of nucleosome positioning, for an improved protein expression.

Thus, in accordance with these particular findings, the present invention provides in a first aspect a method for obtaining the DNA sequence pattern determining the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising the following steps:

a) aligning mononucleosomal DNA sequences of well-positioned nucleosomes of said eukaryotic species, wherein each mononucleosomal DNA sequence is aligned with respect to its central position;

b) obtaining the position-specific weight matrix (PSWM) characteristic for said species, wherein score values for each position (i) and for each k-mer (k) are calculated according to the expression:

$$Score_i^K = log_2 \frac{Freq_i^K}{Freq_{genomic}^K}$$

where $Freq_i^K$ is the frequency of the k-mer at position (i) in the group of aligned sequences and $Freq_{genomic}^K$ is the average genomic frequency of the k-mer;

wherein the term k-mer refers to an oligomer with a number of base pairs from 1 to n/10, wherein n is the length of the mononucleosomal DNA sequences aligned in step a); and wherein each of the elements of the matrix represents the score for each k-mer in said species depending on its position along the mononucleosomal DNA.

In another aspect, the invention relates to a method for tailoring the sequence of a DNA molecule to obtain the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising steps a) and b) as defined under the first aspect of the invention.

In a further aspect, the invention relates to a method for the optimization of a DNA sequence for its expression in a host eukaryotic species based on the modification of its nucleosome positioning, wherein said method comprises steps a) and b) as defined under the first aspect of the invention.

In an additional aspect, the invention relates to a method for nucleosome positioning prediction based on the information in the position-specific weight matrix (PSWM) obtained further to steps a) and b) as defined under the first aspect of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Sequence changes destabilize nucleosome positioning in the *ura4* gene

A) Nucleosomal organization of the *S. pombe ura4* region. MNase-Seq data are taken from Soriano et al. (2013). The *ura4* gene is represented by a pointed rectangle and the coding region is shown in dark grey. Restriction sites for Hind III (H) and Tfi I (T) and the localization of the two hybridization probes (light grey) are indicated. **B)** Chromatin from wild type cells (WT) and from the indicated mutant strains were digested with increasing amounts of MNase (triangles) prior to digestion with Hind III or Tfi I (Shuffled). Samples were electrophoresed, blotted and hybridized to probe 2. Nucleosomes are numbered as in the diagramme above. Brackets indicate regions of increased accessibility to MNase in strains 1_6, 1_3 and 3_4 relative to the wild type. **C)** WT and 1_6 membranes in B) were stripped and rehybridized using probe 1. Bracket indicates the same region in 1_6 cells as shown in B. Controls of naked DNA digested with MNase, and of chromatin incubated without MNase, are shown in Figure 11. Similar controls were carried out for all the remaining MNase experiments.

Figure 2. Sequence changes destabilize nucleosome positioning in non-transcribed regions

A) Nucleosomal organization of two genomic regions of *S. pombe*. Exons (squared and pointed rectangles), introns (thin line) and coding regions (dark grey) of the indicated genes are shown. The sequence underlying nucleosomes shown in grey was randomized and used to replace the wild type versions. Restriction sites for PdmI (P) and NsiI (N) and hybridization probes (light grey bar) are indicated. **B)** MNase analysis from wild type cells (WT) and from the two strains with randomized sequences (Random). Nucleosomes are numbered as in the diagrammes above. The positioning of those corresponding to the randomized sequences (grey) is modified relative to the wild type patterns.

Figure 3. Mononucleosomal DNA encodes portable positioning information

A) White ovals represent nucleosomes across the *ura4* region as shown in Figure 1A. Ovals 2 to 5 indicate regions that were replaced by non-adjacent mononucleosomal sequences from unrelated genomic regions. In nucleosomes 1 and 6, only shadowed regions were replaced by equivalent regions from other mononucleosomal DNA sequences. The restriction site for Psu I (P), and the hybridization probe (light grey bar) are indicated. **B)** The MNase analysis of wild type cells (WT) and of the chimaeric construct, made up of the ectopic sequences, generated a comparable positioning pattern. Nucleosomes are numbered as in the diagrammes above. Nucleosome positioning was lost after individual randomization of the shadowed sequences in A (Random).

Figure 4. Positioning information is dispersed across mononucleosomal DNA

A) Ovals represent nucleosomes across the *ura4* region as in Figure 1A. Shadowed segments indicate the third of the mononucleosomal DNA sequence of nucleosomes 1 to 6 that was replaced by synonymous codons. The replaced sequences were centred on the midposition (dyad) of mononucleosomal (lines) or linker DNA (grey). In the third strain (*int*), the replaced sequences span the two remaining internal regions of each mononucleosomal DNA (black). The restriction site for Hind III and the hybridization probe (light grey bar) are indicated. **B)** MNase analysis of wild type (WT) and the three mutant strains modified in the dyad, linker or internal regions, described in A.

Figure 5. Position-specific weight matrix of nucleosomal signatures

Heatmap representation of the position-specific weight matrix (PSWM) for the indicated four species. X-axis indicates positions relative to the nucleosomal dyad. Y-axis indicates the lod score of the 16 dinucleotides along mononucleosomal DNA calculated as the ratio of their frequency at each position relative to their genomic frequency. Bars on the right represent a colour scale associated to the lod score values.

Figure 6. Engineering nucleosomal positioning on synthetic DNA sequences

A) A+T base composition across an artificial 1 kb fragment incorporating information from nucleosomal signatures of *S. pombe* and *S. cerevisiae*. The predicted positions of nucleosomes are indicated by circles of different grey intensity (see text and Materials and Methods for details on the generation of the sequences). **B)** Genomic regions of the two species where the artificial sequences were inserted (dark grey arrowheads). The restriction sites for Hind III (H), Bsm I (B) and Pag I (P), and the localization of the hybridization probes (light grey bar) are indicated. **C)** The MNase analysis of the *S. pombe* sequence integrated in *S. pombe* generates a regular nucleosomal profile as predicted in A (Seq-Sp/Sp). Insertion of the same sequence in *S. cerevisiae* generates an irregular pattern (Seq-Sp/Sc, bracket). **D)** The *S. cerevisiae* sequence

generates a regular profile after integration in *S. cerevisiae* (Seq-Sc/Sc) but fails to position nucleosomes when integrated in the *S. pombe* genome (Seq-Sc/Sp, bracket).

Figure 7. Engineering nucleosomal positioning on eukaryotic coding DNA sequences

A) Nucleosomal distribution across the *ura4* region of the indicated species (black). MNase-Seq data for *S. octosporus* and *S. japonicus* are taken from Quintales et al. (2015a). The ORF of the three *ura4* genes is shown in dark grey (S.pombe), light grey (S.octosporus) and grey (S.japonicus). Restriction sites for Hind III (H) and Tfi I (T) and the localization of the hybridization probe (light grey bar) are indicated. **B)** MNase analysis of wild type *S. pombe ura4* (left panel). Horizontal lines indicate the same internucleosomal sites as in the diagramme in A. These bands are more diffused (brackets) when the *S. pombe ura4* ORF is replaced by the *S. octosporus* ORF (Native). The original sharp profile is restored when the codons of *S. octosporus ura4* are replaced by synonymous codons with the highest score in the PSWM of *S. pombe* (Remastered). **C)** Replacing the *S. pombe ura4* ORF by the *S. japonicus* ORF abolishes the regular nucleosomal pattern (Native) but it is restored when the *S. japonicus* codons are replaced by synonymous codons as in B (Remastered).

Figure 8. Engineering nucleosomal positioning on prokaryotic genes

A) Nucleosomal distribution across the indicated regions in *S. pombe* and *S. cerevisiae* (black). MNase-Seq data for *S. cerevisiae* are taken from Soriano et al. (2014). Genes are represented by pointed rectangles. Restriction sites for Xho I (X), Pst I (P), Ava II (A), Kpn 2I (K) and Bstx I (B) and the localization of the hybridization probes (light grey bar) are shown. **B)** MNase analysis of wild-type *S. pombe SPBC16G5.03* (left panel). Replacing the *SPBC16G5.03* ORF by the kanamycin ORF (Kan_WT) reduces the intensity and the position of some of the bands. Replacement of the kanamycin codons according to the *S. pombe* nucleosomal signature generates a pattern undistinguishable from that of the *SPBC16G5.03* gene (Kan_Sp). The same sequence is unable to position nucleosomes when it replaces the *YKL007W* ORF in the genome of *S. cerevisiae* (Kan_Sc). **C)** MNase analysis of wild-type *S. cerevisiae YKL007W* (left panel). Replacing the *YKL007W* ORF by the kanamycin ORF modifies the regular nucleosomal profile (Kan_WT). A regular pattern is restored over the kanamycin ORF when its codons are replaced by synonymous codons following the *S. cerevisiae* nucleosomal signature (Kan_Sc). This sequence, however, is unable to generate a regular pattern when it replaces the *SPBC16G5.03* ORF in the genome of *S. pombe* (Kan_Sp).

Figure 9. RNA polymerase II occupancy along wild-type and modified *S. pombe ura4* gene.

A) Diagramme of the *ura4* gene. The transcription start (TSS) and transcription termination (TTS) sites are indicated. The ORF is shown in grey. One of the regions tested by Q-PCR maps to the 3' end of the transcript (light grey bar, Ter) and the other encompasses the TSS (dark grey bar, Prom). These two regions are present in all the strains with modified ORFs studied in this work. **B)** Sonicated chromatin was immunoprecipitated with an anti RNA pol II antibody and the abundance of the Prom and Ter regions was measured by quantitative PCR. Results of all the strains in Figure 1 were normalized to an internal control at the *mat* locus and then referred to the strain harbouring the WT version of the *ura4* ORF. The standard deviation of 3 independent experiments is shown.

Figure 10. Distribution of MNase cutting sites along the *ura4* ORF.

The position of nucleosomes 1 to 7 shown in Figure 1 is represented at the bottom. The ORF of the *ura4* gene is shown in dark grey. The distribution of cutting sites of MNase (vertical lines) is concentrated within internucleosomal linker regions in wild-type cells but it is dispersed along the ORF in the 1_6 and *Shuffled* strains, consistent with the end-labelling mapping shown in Figure 1.

Figure 11. Micrococcal nuclease analysis of wild type and modified *S. pombe* sequences.

A) Chromatin from wild-type cells (WT) and from the indicated mutant strains were digested with increasing amounts of MNase (indicated with triangles, see Materials and Methods) prior to digestion with the appropriate restriction enzymes as in Figure 1. Lanes N contain naked DNA digested with MNase and lanes 0 contain chromatin incubated without MNase under the same conditions as the five samples treated with the enzyme. Brackets at the bottom of the panels indicate the three gel lanes as shown in Figure 1. **B)** Analysis of the intergenic region between the *SPAC22E12.03c* and *ccs1* in Figure 2.

Figure 12. Distribution of MNase cutting sites along orthologous *ura4* ORFs.

The position of nucleosomes 1 to 7 and the internucleosomal lines shown in Figure 7 are represented at the bottom. The distribution of cutting sites of MNase (vertical lines) is concentrated within internucleosomal linker regions in the *S. pombe ura4* ORF. The distribution is slightly more dispersed in the *S. octosporus* native ORF. In the *S. japonicus* native ORF sites are much more dispersed with some clustering at regions not coinciding with those in the *S. pombe* ORF. By contrast, the *S. octosporus* and *S. japonicus* remastered ORFs adopt a strict and regular nucleosomal positioning coinciding with the pattern in *S. pombe* as shown by the end-labelling mapping shown in Figure 7.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

The term "nucleosome" as used herein may refer to the basic structural unit of chromatin. Eukaryotic genomes are organized within the nucleus of cells in a complex formed by DNA and proteins called chromatin. Histones are the most abundant chromatin proteins and form complexes of eight subunits called nucleosomes. The nucleosome core particle is known at crystallographic resolution: ~147 base pairs (bp) of DNA are wrapped in ~1.65 left-handed turns around octamers containing two copies each of four histone types, H2A, H2B, H3, and H4. Core particles are connected by more accessible linker DNA with average linker length varying between species (e.g., 21bp for *S. cerevisiae* or 6 bp for *S.pombe*) or between different cell types within species. The core particle plus linker DNA may be referred as "nucleosome," but this term may also be used for the core particle. (Lieleg et al. 2015, Lantermann et al. 2010).

The term "positioning' or "nucleosome positioning" as used herein refers to the "translational position" of a nucleosome core particle along a DNA sequence, i.e., which bp is at the central dyad position of the core particle. It is a measure of the extent to which a population of nucleosomes resists deviating from its consensus location along the DNA and can be thought of in terms of a single reference point on the nucleosome, like its dyad or central position (Albert, I. et al. 2007; Pugh, B. 2010).

A "well positioned nucleosome" or "translationally well positioned nucleosome" as used herein refers to those where the dyad position of the mononucleosomal DNA corresponds or is about that of the consensus location. Well positioned nucleosomes are often identified by reference to its dyad position. For instance, in Quintales et al. 2015a, well positioned nucleosomes were identified by identification of its dyad position, namely peak positions whose level of occupancy was above the genome average occupancy and the nearest maximum on each direction was at least 120 nucleotides away

The term "occupancy" or "nucleosome occupancy" as used herein refers to the probability of a given base pair to be part of any nucleosome core and amounts to a 147-bp sliding window of translational positioning (Lieleg et al. 2015). For a given position (e.g., peak position), it corresponds to the percentage of a template/cell population with a nucleosome in said position (partially represented by peak height, see "Comparison of different methods and the problem of nucleosome occupancy". For more detailed definitions, see Kaplan et al. 2010). Occupancy may be considered as a measure of the nucleosome density (Pugh, B. 2010).

The term “nucleosomal signature” as used herein may refer to sequence patterns that are found in mononucleosomal DNA sequences and which have been associated to nucleosome positioning.

Quintales et al. 2015a discloses a species-specific nucleosomal signature consisting of an asymmetric base composition pattern (mononucleotide, dinucleotide or trinucleotide pattern) found in the mononucleosomal DNA sequence of well-positioned nucleosomes. More specifically, said species-specific nucleosome signature is obtained by a method comprising:

a) aligning the mononucleosomal DNA sequences of well-positioned nucleosomes, wherein each mononucleosomal DNA sequence is aligned with respect to its central position; and
b) calculating the frequencies of mono-, di- or tri- nucleotides for each position and normalizing to the corresponding genome averages.

The term "identity" as used herein refers to an exact nucleotide-to-nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity”. The “percent identity” of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. Suitable programs for calculating the percent identity or similarity between sequences are well known in the art, such as the NCBI BLAST program, used for example with default parameters (http://www. ncbi. nlm. gov/cgi-bin/BLAST).

Detailed description

In a first aspect, the invention relates to a method for obtaining the DNA sequence pattern determining the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising the following steps:

a) aligning mononucleosomal DNA sequences of well-positioned nucleosomes of said eukaryotic species, wherein each mononucleosomal DNA sequence is aligned with respect to its central position;

b) obtaining the position-specific weight matrix (PSWM) characteristic for said species, wherein score values for each position (i) and for each k-mer (k) are calculated according to the expression:

$$Score_i^K = log_2 \frac{Freq_i^K}{Freq_{genomic}^K}$$

where $Freq_i^K$ is the frequency of the k-mer at position (i) in the group of aligned sequences and $Freq_{genomic}^K$ is the average genomic frequency of the k-mer;

wherein the term k-mer refers to an oligomer with a number of base pairs from 1 to n/10, wherein n is the length of the mononucleosomal DNA sequences aligned in step a); and wherein each of the elements of the matrix represents the score for each k-mer in said species depending on its position along the mononucleosomal DNA.

Said DNA sequence may be any kind of DNA sequence. Illustrative, non-limiting examples include: a non-preexisting artificial sequence (e.g. based on solid-phase DNA synthesis), or a sequence from a virus, from a prokaryotic species or from a eukaryotic species. In a particular embodiment, said DNA sequence is a sequence from a eukaryotic species. Said eukaryotic species can be the same or different from the particular eukaryotic species which characteristic nucleosome positioning is to be obtained. Preferably, it is a sequence from another eukaryotic organism, more preferably, from another species or strain.

The term "mononucleosomal DNA" as used herein may refer to the 147 bp of core DNA that are wrapped around the histone octamers or to the 147 bp of core DNA plus the linker sequence. The size of the linker DNA may differ depending on the species, typically ranging from 2 to 50 bp, for instance it has a length of 21bp for *S. cerevisae* and 6 bp for *S.pombe.* Preferably, it refers to the 147 bp of core DNA.

Mononucleosomal DNA sequences of well-positioned nucleosomes are aligned in step a) of the method of the invention. Well-positioned nucleosomes in a genomic sequence are identified thanks to the generation of nucleosome occupancy maps, often by reference to its dyad or central position (see above the definition of "well-positioned nucleosomes").

Nucleosome occupancy maps may have different formats, for example be based on "nucleosome occupancy" i.e., the signal at each bp is contributed by any nucleosome covering this bp, or "nucleosome dyad occupancy/density," i.e., only base pairs at nucleosome dyads contribute, the latter presenting more distinct peaks (Lieleg et al. 2015, Zhang and Pugh 2011a).

Many methods are known in the art for the generation of nucleosome occupancy maps. These typically include the use of enzymes, like nucleases or methylases, or chemicals to distinguish nucleosomal from non-nucleosomal DNA. Illustrative non limiting examples are described in Lieleg et al. 2015 and Teif 2015, and include: mapping by micrococcal nuclease (MNase), MNase-array, MNase-sequencing, MNase-sequencing including a step of chromatin immunoprecipitation after MNase digestion (MNase-ChIP-seq), MNase-ChIP-qPCR, transposase (ATAC-seq), CpG methyltransferase (NOME-seq), hydroxyl-radical-seq, MNase-exoIII-seq, formaldehyde-assisted isolation of regulatory elements (FAIRE), particle spectrum analysis, etc.

In a particular embodiment, nucleosome occupancy maps are generated by a method comprising isolation and sequencing of the mononucleosomal DNA. Mononucleosomal DNA isolation is generally performed by a method comprising mononuclease digestion of chromatin. This is typically performed until the generation of a 80:20 ratio of mononucleosomes to dinucleosomes as described by Lantermann et al. 2010. Subsequently, the entire mononucleosomal fragment may be sequenced or only its ends, either both ends together (paired end) or just one end (single end). Methods for DNA sequencing are well known in the art. These include Sanger sequencing and next-generation sequencing. A review of next-generation DNA sequencing methods is provided for instance in Mardis E.R. 2008, Metzker ML 2010 or Goodwin S et al. 2016.

Generation of genomic nucleosome occupancy maps typically involves mapping the mononucleosomal DNA sequence reads to the corresponding reference genomes. A person skilled in the art will know how to identify the appropriate reference genomes. For instance, the reference genomes of the wild type yeast species used in the examples were *S. pombe* (ASM294v2.20, assembly 13 August 2013) from PomBase; *S. octosporus* (SO6, assembly 7 June 2012) and *S. japonicus* (SJ5, assembly 7 June 2012) from the Broad Institute *Schizosaccharomyces* group Database, and *S. cerevisae strain S288C* (R64-1-1, assembly 3 February 2011) from the *Saccharomyces* Genome Database. The mathematical treatment of the obtained signal (e.g. smoothing using a wavelet process) will provide the final occupancy map.

A great number of computational tools to analyse nucleosome positioning data have been described. Teif 2015 provides a systematic presentation of existing bioinformatic approaches and resources (in particular in pages 2-4 and Table 1). Nucleosome positioning may be based on nucleosome calling. Generic peak calling programs usually used for ChIP-seq, such as MACs or HOMER, are generally found not optimal for nucleosome position calling. Yet, the

basic idea behind nucleosome position calls from the experimental data is the same: one has to detect enriched peaks of size around 147bp (e.g. TemplateFilter, NPC, nucleR, NOrMAL, PING/ PING2, MLM, NucDe, NucleoFinder, ChIPseqR, NSeq, NucHunter, iNPS and PuFFIN). One alternative is not calling nucleosome positions at all, and instead operating with the continuous nucleosome occupancy profile, defining regions of cell type/state specific differential occupancy (e.g. DANPOS/DANPOS2, DiNuP, and NUCwave).

Accordingly, for instance the automatic generation of nucleosome occupancy maps may be provided by the NUCwave bioinformatic tool, as described in Quintales et al. 2015b. In a preferred embodiment, further to the sequencing of mononucleosomal DNA nucleosome, occupancy maps are generated using the NUCwave algorithm. See for instance, “preparation of mononucleosomal DNA and next generation sequencing” in the Examples.

The central position of well-positioned nucleosomes may be identified for instance by determining in the final occupancy map peak positions whose level of occupancy is above the genome average occupancy and defining the nearest maximum on each direction. A person skilled in the art will know that a higher distance between peaks would result in a more accurate positioning. The distance between peaks may be for example of at least about 90, at least about 100, at least about 110, at least about 120 or at least about 150 nucleotides away. In a preferred embodiment the distance between peaks is of at least 120 nucleotides away. See Quintales et al. 2015a, section 3.5.

A person skilled in the art will know how to find the appropriate amount of mononucleosomal DNA sequences with well-positioned nucleosomes to be aligned in step a). The number is generally such that will enable the signal versus background noise to obtain a statistically significant profile. The number of aligned mononucleosomal DNA sequences in step a) is not limited, and is typically of at least 100, preferably of at least 200, 500, 1000, preferably of at least 5000, more preferably of at least 10.000. The number of aligned sequences will typically depend on the size of the genome of the particular eukaryotic species, the number of sequences to be aligned generally increasing with the size of the genome. For instance, for the obtaining of the PSWM in the yeast species used in the examples, 38154, 46120, 27024 and 34526 mononucleosomal DNA sequences of *S. pombe*, *S. octosporus*, *S. japonicus* and *S. cerevisiae*, respectively, were aligned to their central (dyad) position (see also Quintales et al. 2015a). In a preferred embodiment, the number of mononucleosomal DNA sequences aligned in step a) is from around 20.000 to around 50.000, more preferably from around 30.000 to around 45.000.

For the obtaining of the position-specific weight matrix (PSWM) characteristic for said eukaryotic species, said k-mer in step b) may have different lengths which size is limited by the length of the mononucleosomal DNA sequence. Typically, the mononucleosomal DNA sequence has a length of around 150 bp and often the k-mer has a number of base pairs which is about 1 to n/10, wherein n is the length of the mononucleosomal DNA sequences aligned in step a). Preferably, the k-mer has a length of 1 to 15 nucleotides, preferably from 1 to 10 nucleotides, more preferably from 1 to 5 nucleotides. In a preferred embodiment, said k-mer in step b) has one, two or three nucleotides.

In a more preferred embodiment, said k-mer in step b) is a dinucleotide (NN) and the score values for each position (i) and for each dinucleotide (NN) are calculated according to the expression:

$$Score_i^{NN} = log_2 \frac{Freq_i^{NN}}{Freq_{genomic}^{NN}}$$

where $Freq_i^{NN}$ is the frequency of the NN dinucleotide at position (i) in the group of aligned sequences and $Freq_{genomic}^{NN}$ is the average genomic frequency of the same dinucleotide,

wherein each of the elements of the matrix represents the score for each dinucleotide in said species depending on its position along the mononucleosomal DNA.

It is noted that when the k-mer is a dinucleotide (2-mer) for each (i) position said dinucleotide can be AA, TT, AT, TA, CC, GG, CG, GC, AC, GT, AG, CT, TG, CA, TC or GA (see figure 5 for illustration purposes).

The eukaryotic species which nucleosome positioning is to be obtained may be any eukaryotic species and includes without limitation, fungi, protozoa, and mammalian species. Table 1 of Teif 2015 provides a summary of available nucleosome positioning datasets for diverse eukaryotic species. In one embodiment, said eukaryotic species is a mammalian species. Preferably, it is selected from a human, companion animal, non-domestic livestock or zoo animal. For example, the subject may be selected from a human, mouse, rat, dog, cat, cow, pig, sheep, horse, bear, and so on. In a preferred embodiment, said mammalian species is a human (*Homo sapiens*).

In another embodiment said eukaryotic species is a fungi species, preferably from the Ascomycota phylum. In a preferred embodiment, said eukaryotic species is a yeast. As used herein, the term “yeast” includes not only yeast in a strict taxonomic sense, ie, unicellular organisms, but also multicellular fungi like yeasts or filamentous fungi. Examples of species are

Kluyveromyces lactis, Schizosaccharomyces pombe, Ustilago maydis, Saccharomyces cerevisiae and *Pichia pastoris*. Other yeasts species encompassed are *Neurospora crassa, Aspergillus niger, Aspergillus nidulans, Aspergillus sojae, Aspergillus oryzae, Candida tropicalis*, and *Hansenula polymorpha*. In a particularly preferred embodiment, said yeast species is selected from the group consisting of *Schizosaccharomyces pombe*, *Schizosaccharomyces octosporus*, *Schizosaccharomyces japonicus* and *Saccharomyces cerevisiae*.

In another aspect, the invention relates to a method for tailoring the sequence of a DNA molecule to obtain the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising steps a) and b) as described under the first aspect of the invention.

The DNA molecule to be tailored may comprise or consist of coding and/or non-coding sequences. In a particular embodiment, said DNA molecule sequence comprises or consists of a coding DNA sequence. The inventors have devised a method based on combining the information on the nucleosomal signatures with the degenerated positions of synonymous codons, and selecting for each amino acid position in the encoded polypeptide the codon which better reproduces the pattern in the nucleosomal signature of a particular eukaryotic species. Accordingly, in a preferred embodiment, said DNA molecule sequence comprises or consists of a coding DNA sequence and the method for tailoring the sequence of a DNA molecule to obtain the nucleosome positioning characteristic of a particular eukaryotic species of the invention further comprises:

i. determining for each codon in the DNA coding sequence the synonymous codon with the highest score in the PSWM of step b); and
ii. replacing each codon in said DNA coding sequence for the synonymous codon determined in (i).

The term “synonymous codon” as used herein refers to a DNA triplet or codon codifying for the same amino acid residue. The genetic code has redundancy but no ambiguity. For example, although codons GAA and GAG both specify glutamic acid (redundancy), neither of them specifies any other amino acid (no ambiguity). The codons encoding one amino acid may differ in any of their three positions. For example, the amino acid leucine is specified by YUR or CUN (UUA, UUG, CUU, CUC, CUA, or CUG) codons (difference in the first or third position indicated using IUPAC notation), while the amino acid serine is specified by UCN or AGY (UCA, UCG, UCC, UCU, AGU, or AGC) codons (difference in the first, second, or third position). The codon table below provides the full correlation.

Table 1.- RNA codon table

Amino acid	Codons	Compressed	Amino acid	Codons	Compressed
Ala/A	GCU, GCC, GCA, GCG	GCN	Leu/L	UUA, UUG, CUU, CUC, CUA, CUG	YUR, CUN
Arg/R	CGU, CGC, CGA, CGG, AGA, AGG	CGN, MGR	Lys/K	AAA, AAG	AAR
Asn/N	AAU, AAC	AAY	Met/M	AUG	
Asp/D	GAU, GAC	GAY	Phe/F	UUU, UUC	UUY
Cys/C	UGU, UGC	UGY	Pro/P	CCU, CCC, CCA, CCG	CCN
Gln/Q	CAA, CAG	CAR	Ser/S	UCU, UCC, UCA, UCG, AGU, AGC	UCN, AGY
Glu/E	GAA, GAG	GAR	Thr/T	ACU, ACC, ACA, ACG	ACN
Gly/G	GGU, GGC, GGA, GGG	GGN	Trp/W	UGG	
His/H	CAU, CAC	CAY	Tyr/Y	UAU, UAC	UAY
Ile/I	AUU, AUC, AUA	AUH	Val/V	GUU, GUC, GUA, GUG	GUN
START	AUG		STOP	UAA, UGA, UAG	UAR, URA

Compressed using IUPAC notation. Source: https://en.wikipedia.org/wiki/Genetic_code

In another particular embodiment, said DNA molecule sequence comprises or consists of a non-coding sequence. The term “non-coding DNA sequence” as used herein refers to components of an organism's DNA that do not encode protein sequences. Some non-coding DNA is transcribed into functional non-coding RNA molecules (e.g. transfer RNA, ribosomal RNA, and regulatory RNAs). Other functions of non-coding DNA include the transcriptional and translational regulation of protein-coding sequences, scaffold attachment regions, origins of DNA replication, centromeres and telomeres. Examples of sequences regulating the transcription of protein-coding sequences include operators, enhancers, silencers, promoters and insulators.

In a preferred embodiment, said DNA molecule sequence comprises or consists of a non-coding DNA sequence and the method for tailoring the sequence of a DNA molecule to obtain the nucleosome positioning characteristic of a particular eukaryotic species of the invention further comprises:

i. identifying the three possible reading frames of trinucleotides in said non-coding DNA sequence;
ii. determining the corresponding codon for each trinucleotide in each of the three possible reading frames;
iii. determining for each codon in each of the three possible reading frames the synonymous codon with the highest score in the PSWM of step b) and replacing it by said synonymous codon; and
iv. selecting from the modified sequences for each of the three possible reading frames obtained in (iii) the one with the highest score in the PSWM.

Alternatively, said method further comprises:

i. submitting said DNA sequence to random single point mutation;
ii. selecting the resulting sequence if the mutated k-mer has a score for the position (i) as determined in the PSWM of step b) higher than in the previous sequence;.

wherein steps i) and ii) are repeated for several cycles. A person skilled in the art would know when to arrest this mutation and selection process. In a particular embodiment, the DNA sequence would have a high score in the PSWM while maintaining the functionality of the non-coding sequence.

The process defined in steps i) and ii) above may for instance be carried out in parallel to several distinct initial DNA sequences until the average percentage of identity between the newly generated sequences is not higher than the average percentage of identity between individual genomic mononucleosomal sequences of said species. In a preferred embodiment, steps i) and ii) are repeated until the average percentage of identity between the newly generated sequences is about 25%. Another option would be to repeat steps i) and ii) until the mutated sequence has an identity with the sequence presenting the highest possible score in the PSWM for each (i) position which is not higher than 95%, preferably not higher than 90%, 85%, 80%, 75% or 70%.

In another aspect, the invention relates to a method for the optimization of a DNA sequence for its expression in a host eukaryotic species based on the modification of its nucleosome positioning, wherein said method comprises steps a) and b) as defined under the first aspect of the invention.

A better nucleosome positioning may be associated with an improved or a decreased protein expression. A person skilled in the art will know how to use the information provided by the PSWM characteristic from a particular eukaryotic species obtained as described herein for optimizing a DNA sequence for improved protein expression in said eukaryotic species. The modified sequences may comprise or consist of coding and/or non-coding sequences.

It has been reported by other authors that strong promoters present a low nucleosome positioning which may facilitate RNA polymerase access (Shivaswamy S et al. 2008, and Badis G. et al. 2008). In this scenario, the information of the PSWM characteristic from a particular eukaryotic species obtained as described herein may be used to modify the sequence of a DNA

molecule to differ from the characteristic positioning of said eukaryotic species in order to decrease nucleosome positioning.

In another particular scenario, better nucleosome positioning results in improved protein expression. Accordingly, in a particular embodiment, the method for the optimization of a DNA sequence for its expression in a host eukaryotic species comprises the tailoring of said DNA sequence to obtain the nucleosome positioning characteristic of said host eukaryotic species by a method as described above. In a preferred embodiment, said DNA sequence consists of or comprises a coding sequence.

Several methods for sequence optimization for improved protein expression are known in the art and specifically include codon optimization methods, also referred as ORF engineering methods. Gustaffsson et al. 2012 is a review on gene engineering for increasing protein expression yields and provides that variables for gene engineering can broadly be classified as local or global variables. Local variables are defined as those variables which are dependent on local sequence patterns within the gene. In general, local variables affect protein expression at the level of the gene DNA or mRNA. Examples are RNA degradation motifs and mRNA secondary structures. In contrast, global variables are measures of aggregate gene features. Examples include codon usage frequency and GC%.

For instance, there are a number of local variables that are considered to be deleterious for heterologous protein expression and many gene design algorithms seek to exclude these from gene sequences by using synonymous codons to encode the same amino acid sequence. Local variables that are generally considered deleterious include but are not limited to RNase sites, DNA recombination sites, transcriptional terminators, and transcription factor recognition sequences. Some motifs depend on the host organism: cryptic splicing and internal polyA signals could affect genes to be expressed in eukaryotic hosts. An Example of multiparameter codon optimization method is provided in Fath et al. 2011. More specifically, the various candidate genes' coding regions were optimized taking the following sequence-based parameters into account: (i) Codon choice, (ii) increase in GC-content, (iii) avoiding UpA- and introducing CpG-dinucleotides, (iv) removing destabilizing RNA elements, (v) removing cryptic splice-sites, (vi) avoiding intragenic poly(A)-sites, (vii) removing direct repeats, (viii) avoiding RNA secondary structures, and (ix) deleting internal ribosomal entry sites.

In a particular embodiment, the information in the position-specific weight matrix (PSWM) for species-specific nucleosome positioning obtained by the method of the invention is used in

combination with known local or global variables in an algorithm for the optimization of a DNA sequence to obtain an increased protein expression in an eukaryotic host.

In a further aspect, the invention relates to a method for nucleosome positioning prediction wherein said method comprises steps a) and b) as defined under the first aspect of the invention. More specifically, the present invention provides a method that enables to predict nucleosome positioning for a particular DNA sequence based on the information in the position-specific weight matrix (PSWM) obtained in step b), by comparing the scoring for each (i) position in the PSWM for said DNA sequence to the highest possible score.

The methods of the present invention might be implemented by a computer. Therefore, a further aspect of the invention refers to a computer implemented method, wherein the method is any of the methods disclosed herein or any combination therefore.

It is noted that any computer program capable of implementing any of the methods of the present invention or used to implement any of these methods or any combination therefore, also forms part of the present invention.
It is also noted that any device or apparatus comprising means for carrying out the steps of any of the methods of the present invention or any combination therefore, or carrying a computer program capable of, or for implementing any of the methods of the present invention or any combination therefore, is included as forming part of the present specification.

The present invention further relates to a computer-readable storage medium having stored thereon a computer program of the invention or the results of any of the methods of the invention. As used herein, “a computer readable medium” can be any apparatus that may include, store, communicate, propagate, or transport the results of the determination of the method of the invention. The medium can be an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system (or apparatus or device) or a propagation medium.

It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method, and use of the invention described herein. It will be understood that particular embodiments described herein are shown by way of illustration and not as limitations of the invention. The principal features of this invention can be employed in various embodiments without departing from the scope of the invention. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific procedures described herein. Such equivalents are considered to be within the scope of this invention and are covered by the claims.

All publications and patent applications mentioned in the specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The use of the word "a" or "an" may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one". The use of the term “another” may also refer to one or more. The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive.

As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. The term “comprises” also encompasses and expressly discloses the terms “consists of” and “consists essentially of”. As used herein, the phrase "consisting essentially of” limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristic(s) of the claimed invention. As used herein, the phrase "consisting of excludes any element, step, or ingredient not specified in the claim except for, e.g., impurities ordinarily associated with the element or limitation.

The term "or combinations thereof” as used herein refers to all permutations and combinations of the listed items preceding the term. For example, "A, B, C, or combinations thereof” is intended to include at least one of: A, B, C, AB, AC, BC, or ABC, and if order is important in a particular context, also BA, CA, CB, CBA, BCA, ACB, BAC, or CAB. Continuing with this example, expressly included are combinations that contain repeats of one or more item or term, such as BB, AAA, AB, BBC, AAABCCCC, CBBAAA, CABABB, and so forth. The skilled artisan will understand that typically there is no limit on the number of items or terms in any combination, unless otherwise apparent from the context.

As used herein, words of approximation such as, without limitation, "about", "around”, “approximately” refers to a condition that when so modified is understood to not necessarily be absolute or perfect but would be considered close enough to those of ordinary skill in the art to warrant designating the condition as being present. The extent to which the description may

vary will depend on how great a change can be instituted and still have one of ordinary skilled in the art recognize the modified feature as still having the required characteristics and capabilities of the unmodified feature. In general, but subject to the preceding discussion, a numerical value herein that is modified by a word of approximation such as "about" may vary from the stated value by ±1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15%. Preferably the term “about” means exactly the indicated value (± 0%).

The following examples serve to illustrate the present invention and should not be construed as limiting the scope thereof.

EXAMPLES

I. MATERIALS AND METHODS

Wild-type and modified DNA sequences

A) Sequences corresponding to Figure 1. Wild-type and modified versions of *Schizosaccharomyces pombe ura4* ORF. Initiation and termination codons (ATG and TAA) are underlined. Sequences harbouring synonymous codons relative to the wild-type ORF are indicated in **bold**. All ORFs have the same size (795 bp) and encode the same Ura4 protein.

ura4 WT (SEQ ID NO:1)

ATGGATGCTAGAGTATTTCAAAGCTATTCAGCTAGAGCTGAGGGGATGAAAAATCCCATTG
CCAAGGAATTGTTGGCTTTGATGGAAGAAAAGCAAAGCAACTTGTCAGTCGCGGTCGATTT
GACGAAGAAATCCGAAATCTTAGAATTGGTAGATAAAATTGGACCCTATGTCTGTGTTATCA
AGACACATATTGACGTTGTCGAGGATTTCGACCAGGATATGGTAGAAAAACTGGTGGCCTT
AGGTAAAAAGCATCGTTTTCTTATCTTTGAGGATCGCAAATTCGCAGACATTGGAAATACCG
TCAAGCTACAATATGCATCTGGTGTGTACAAAATTGCTTCTTGGGCTCATATCACAAATTGC
CATACAGTGCCAGGCGAGGGTATTATACAAGGCCTCAAAGAAGTTGGTTTACCTTTGGGAC
GTGGTCTCTTGCTTTTGGCTGAAATGTCTTCCAAAGGCTCTTTGGCTACTGGTTCCTACACA
GAGAAAACCTTAGAATGGTTTGAGAAGCATACCGATTTTTGCTTTGGCTTTATAGCTGGTCG
TCGATTTCCTAACCTTCAAAGCGACTACATAACTATGTCCCCTGGTATCGGCTTGGATGTTA
AAGGAGACGGGCTGGGACAGCAATATCGTACTCCTGAAGAAGTGATTGTAAACTGCGGTA
GCGATATCATCATTGTTGGTCGTGGAGTCTATGGAGCTGGTCGTAATCCTGTTGTCGAAGC
CAAGAGATATAGAGAAGCTGGTTGGAAGGCATATCAGCAAAGACTTTCTCAGCATTAA

ura4 1_6 (SEQ ID NO:2)

ATG**GACGCGCGCGTGTTCCAGTCATACAGCGCGCGCGCGGAAGGTATGAAGAACCCGA**
TCGCTAAAGAGCTACTAGCGCTAATGGAGGAGAAACAGTCAAATCTAAGCGTAGCTGTA
GACCTAACTAAAAAGAGTGAGATACTCGAGCTAGTCGACAAGATCGGGCCATACGTATG
CGTCATTAAAACGCACATCGATGTCGTAGAAGACTTTGATCAAGACATGGTCGAGAAGTT
AGTAGCACTCGGCAAGAAACACAGGTTCTTGATATTCGAAGACAGAAAGTTTGCGGATA
TCGGCAACACTGTTAAATTGCAGTACGCGAGCGGCGTATATAAGATCGCGAGCTGGGCG
CACATAACGAACTGTCACACCGTACCGGGTGAAGGCATCATCCAGGGATTAAAGGAGGT
CGGCCTGCCGCTAGGCAGGGGCTTACTATTGCTAGCAGAGATGAGCAGTAAGGGAAGC

CTAGCGACGGGCAGTTATACGGAAAAGACACTCGAGTGGTTCGAAAAACACACTGACTT
CTGTTTCGGATTCATCGCGGGCAGGAGGTTCCCGAATTTGCAGTCAGATTATATCACGAT
GAGTCCGGGCATAGGACTAGACGTCAAGGGCGATGGATTAGGCCAACAGTACAGGACG
CCGGAGGAGGTAATCGTCAATTGTGGCTCAGACATTATTATCGTCGGCAGGGGCGTATA
CGGCGCCGGCAGGAACCCGGTCGTAGAGGCTAAACGCTACCGCGAGGCCGGCTGGAA
AGCGTACCAACAGCGCTTGAGCCAACACTAA

ura4 1_3 (SEQ ID NO:3)

ATG**GACGCGCGCGTGTTCCAGTCATACAGCGCGCGCGCGGAAGGTATGAAGAACCCGA**
TCGCTAAAGAGCTACTAGCGCTAATGGAGGAGAAACAGTCAAATCTAAGCGTAGCTGTA
GACCTAACTAAAAAGAGTGAGATACTCGAGCTAGTCGACAAGATCGGGCCATACGTATG
CGTCATTAAAACGCACATCGATGTCGTAGAAGACTTTGATCAAGACATGGTCGAGAAGTT
AGTAGCACTCGGCAAGAAACACAGGTTCTTGATATTCGAAGACAGAAAGTTTGCGGATA
TCGGCAACACTGTTAAATTGCAGTACGCGAGCGGCGTATATAAGATCGCGAGCTGGGCG
CACATAACGAACTGTCACACCGTACCGGGTGAAGGCATCATACAAGGCCTCAAAGAAGTT
GGTTTACCTTTGGGACGTGGTCTCTTGCTTTTGGCTGAAATGTCTTCCAAAGGCTCTTTGG
CTACTGGTTCCTACACAGAGAAAACCTTAGAATGGTTTGAGAAGCATACCGATTTTTGCTTT
GGCTTTATAGCTGGTCGTCGATTTCCTAACCTTCAAAGCGACTACATAACTATGTCCCCTG
GTATCGGCTTGGATGTTAAAGGAGACGGGCTGGGACAGCAATATCGTACTCCTGAAGAAG
TGATTGTAAACTGCGGTAGCGATATCATCATTGTTGGTCGTGGAGTCTATGGAGCTGGTCG
TAATCCTGTTGTCGAAGCCAAGAGATATAGAGAAGCTGGTTGGAAGGCATATCAGCAAAGA
CTTTCTCAGCAT**TAA**

ura4 3_4 (SEQ ID NO:4)

ATGGATGCTAGAGTATTTCAAAGCTATTCAGCTAGAGCTGAGGGGATGAAAAATCCCATTG
CCAAGGAATTGTTGGCTTTGATGGAAGAAAAGCAAAGCAACTTGTCAGTCGCGGTCGATTT
GACGAAGAAATCCGAAATCTTAGAATTGGTAGATAAAATTGGACCCTATGTCTGTGTTATCA
AGACACATATTGACGTTGTCGAGGATTTCGACCAGGATATGGTAGAAAAACTGGTGGCCTT
AGGTAAAA**AACACAGGTTCTTGATATTCGAAGACAGAAAGTTTGCGGATATCGGCAACAC**
TGTTAAATTGCAGTACGCGAGCGGCGTATATAAGATCGCGAGCTGGGCGCACATAACGA
ACTGTCACACCGTACCGGGTGAAGGCATCATCCAGGGATTAAAGGAGGTCGGCCTGCC
GCTAGGCAGGGGCTTACTATTGCTAGCAGAGATGAGCAGTAAGGGAAGCCTAGCGACG
GGCAGTTATACGGAAAAGACACTCGAGTGGTTCGAAAAACACACTGACTTCTGTTTCGG
ATTCATCGCTGGTCGTCGATTTCCTAACCTTCAAAGCGACTACATAACTATGTCCCCTGGTA
TCGGCTTGGATGTTAAAGGAGACGGGCTGGGACAGCAATATCGTACTCCTGAAGAAGTGA
TTGTAAACTGCGGTAGCGATATCATCATTGTTGGTCGTGGAGTCTATGGAGCTGGTCGTAA
TCCTGTTGTCGAAGCCAAGAGATATAGAGAAGCTGGTTGGAAGGCATATCAGCAAAGACTT
TCTCAGCAT**TAA**

ura4 Shuffled (SEQ ID NO:5)

ATG**GACGCGCGTGTCTTCCAGTCCTACTCTGCCCGAGCAGAAGGCATGAAGAACCCAAT**
CGCCAAAGAGCTCCTGGCCCTAATGGAGGAAAAGCAAAGCAATCTCTCAGTGGCTGTGG
ACCTGACGAAGAAGTCAGAAATCTTAGAGCTTGTTGACAAAATTGGCCCCTACGTTTGTG
TTATTAAGACCCATATAGACGTCGTGGAAGACTTTGATCAAGATATGGTAGAAAAACTTG
TTGCATTGGGTAAGAAACATCGTTTCCTTATCTTTGAGGATCGTAAATTTGCTGATATCGG
AAACACTGTGAAATTACAGTACGCAAGCGGCGTTTATAAAATTGCTAGCTGGGCTCATAT
TACCAACTGCCATACCGTCCCTGGAGAAGGGATAATCCAGGGACTTAAGGAAGTAGGAT
TACCTTTAGGAAGAGGGTTGTTGTTGTTGGCTGAAATGTCCAGCAAAGGATCCTTGGCTA
CTGGTTCCTATACTGAAAAAACATTGGAATGGTTTGAGAAGCATACAGATTTTTGCTTTGG
CTTTATTGCTGGAAGACGCTTTCCCAATTTGCAGTCTGATTATATAACAATGTCTCCTGGC
ATTGGTTTGGATGTAAAGGGTGATGGTTTGGGTCAACAATATCGTACACCTGAGGAAGTT
ATCGTAAATTGCGGTTCTGATATTATCATTGTCGGTAGAGGTGTCTATGGTGCTGGTAGA
AATCCTGTCGTCGAAGCTAAACGTTATAGAGAAGCTGGTTGGAAAGCTTATCAACAACGT
TTGTCTCAACATTAA

B) Sequences corresponding to Figure 2. Wild-type and randomized sequences underlying nucleosomes 5-7 of Intergenic region SPAC22E12.03c-ccs1 and nucleosomes 3-4 of Intergenic region SPAC6F6.11c-SPAC6F6.12 of S. pombe. The genomic position of the first and last nucleotide of each sequence and the genes flanking the two intergenic regions are indicated.

Nucleosomes 5-7

Chromosome I: Positions 5024357 to 5024796 (Intergenic region SPAC22E12.03c-ccs1)

Wild type sequence (SEQ ID NO:6)

ATATATACGAATCACTAAAATTGTGAGAGTACTGTAATCACTAAACAGAACGTTCAAATTCTC
CATAATATAAGAAAAATAATACTTAGAATTTAGATATTCCTACATCTCTCTATCGCCCATAGC
ATTTATTTTATTTTTTTAAATCAAACGAAGGAGTTGTACACTAACAATTACAGCTAAATGGTC
TTTAAAAAATTAATGAAAAATAAATAATAATACAATCTCTATGTTCATTTATAAACCAACAGAA
AGAAAAGCCATTTAACTCGCTTGGAATAAAAAACAAATCTTCCAAAAAAATTTAAAAAAATAA
ATAATAATAAATAGAAAAATCCAACATTCATTGTCATTTCTAATAAAATCAATGATATTCTTTA
AAAATTGACGTTCATATTTTGCATTTTTCAATTCAATATACCATTAACTTTCCTTCTGAC

Randomized sequence (SEQ ID NO:7)

TTCATTCTATCGTAATTATCAGAGACTCTATAATCGTTTATAATCAAGAACTAATACTACAAA
TACAGAAGAAATATCAATTATTTCGAAGTATATATTTTAATATCTTTCTAGCCACACTCCATC
AAAAAAGTTAGCTAAATTTACTAAGTATTAACTTAAATACACATCTTCAAATATTAAACTTTAT
GGAAAAATATACACAAATATAAAAACAGATAATTTTAAAAAAAGTAGAAATTTTAATAAATTAT
ACTGACCTCAATTTATTTGACCTAAGTAAAGACTCCGAACTATAATATACACATAAATTACCA
ACTCATTATCTTTTTTTAAGTACAAATTCACGTTTCCCAATAATAAAACTGTCAACAAAATATG
GCACTTGATGTAAGTAAACCTTGAAAACTGAGTATAAAATATCCATCTAGATTAGTTAA

Nucleosomes 3-4

Chromosome I: Positions 2753585 to 2753879 (Intergenic region SPAC6F6.11c-SPAC6F6.12)

Wild type sequence (SEQ ID NO:8)

GATAATCAACTTAATTTTTTGCGACAAAACTAATTATGAAGCACATTCTTTTGAATCAATTTAT
CCAAAAAAGTAATAAATTCAACAGTGACAAGATGTAAGCATTATACGGAATACAACGACTCG
AAATACATGGAATTGTATAACTTAATCACTCATTTAAACTTTTACAGAAATCTTCGGCTGAAA
TATACGGGGTAAAATAGGGCTTTTAATTGTAAAGCTTAGAGCCTGAATAAGTATGCTTTCCT
TTATTTTAACATTCCTCAATTTGTTTTCAAGTACTTCACATACC

Randomized sequence (SEQ ID NO:9)

GGTACTAATATCAACATTTAAAATGTGATTTGGTTCTTTATCTTATGTAACCTTAGAAATAATT
TTTTATACGACAAATAACTTGTGGCCTTGTAACATCGGTATAAACAATTTCATTAATATTAAA
TATTAATTATACAGAGTACAGTTCAAATCTAAAGTAGTTCGCAACTCTACTCTCGCATCCTCT
CGATAAACAAACATAGAAGTAATACAAAAGCCAGAAATTGCTCTAAAAATGAATTTTGTACC
GTAGGTATCCTTTGTGTTACCGCATATAAATACTGATACGAAG

C) Sequences corresponding to Figure 3. Sequence and genomic position of the six fragments used to assemble the chimaeric sequence and the sequence resulting from the randomization of the six individual sequences.

Nucleosome 1 (SEQ ID NO:10)
Chromosome III: Positions 2236704 to 2236793 (reverse complementary)
TATAGTATCAAATTTTTGTGTCTTAAGCTTTTTATGCAGATGCAGTTTCGTTGCATCAATTTC
ATGCATAACCTTTCCCCCATTTATTTC

Nucleosome 2 (SEQ ID NO:11)
Chromosome III: Positions 844797 to 844949
GATACAAGGACTCTAAAAATCGCCAGATACAGAATGCGTCAAAACCTTGAGAAGGAGTTAA
AGGCTTTGTCTAAGGACTCGGAAACCAATGAAGAGCAGGAGAGGAAATTTTGGCTCACTAA
ATTACAAATTGCTGTTGAAGATACATTGGAC

Nucleosome 3 (SEQ ID NO:12)
Chromosome II: Position 3119782 to 3119934 (reverse complementary)
GTTTTTCAATACTAGGAGATCGCTATCCAGAATGCAGACCTTATCAAACTGTGTTTGCTCAA
ATACACTGAGTTTTGTGAACATGTATTGATAACGATAACTGCCCATGCCAATAGACAAACCA
CTTTCAATTAAGTCTTCAGTTTTAATTTG

Nucleosome 4 (SEQ ID NO:13)
Chromosome I: Position 2522247 to 2522399
TATTTTTCTCAAGGCGATTTTTCGAACTTAGATGTACTACGAGATTTTTTTCATGTACTCACT
AGGAAGGAGCAAGATGTCCACATTGCTCTTAACATCAACTTGAAAAGTTTTTTGTGCGAATG
GAGACAGAATGCTCTTGTTCTACTAAAA

Nucleosome 5 (SEQ ID NO:14)
Chromosome I: Position 1502152 to1502304
AATACCAATAAGTCACACCAAAGGTCTTTTTCAATTCCAAAAGCAACGAAGGATTCTCAAAC
TCCGTCGGAAAATAGTGCTGCAACTTTAAAACAAGCTGCAATTGATGCGTATTCTCAAATTC
CTGTCATTCCTTTTTTTCTTCCCAGTGAT

Nucleosome 6 (SEQ ID NO:15)
Chromosome I: Position 807303 to 807389
GTTCCATCTACTAGAACATTTAGTCAGCTTCTAAACACAAAGCAGGCTTCTTTCCTTGATTTT
CTTTCAAAATGTTTGAAATGGGAT

Chimaeric sequence (SEQ ID NO:16)
GAAATAAATGGGGGAAAGGTTATGCATGAAATTGATGCAACGAAACTGCATCTGCATAAAA
AGCTTAAGACACAAAAATTTGATACTATAGATACAAGGACTCTAAAAATCGCCAGATACAGA
ATGCGTCAAAACCTTGAGAAGGAGTTAAAGGCTTTGTCTAAGGACTCGGAAACCAATGAAG
AGCAGGAGAGGAAATTTTGGCTCACTAAATTACAAATTGCTGTTGAAGATACATTGGACCAA
ATTAAAACTGAAGACTTAATTGAAAGTGGTTTGTCTATTGGCATGGGCAGTTATCGTTATCA
ATACATGTTCACAAAACTCAGTGTATTTGAGCAAACACAGTTTGATAAGGTCTGCATTCTGG
ATAGCGATCTCCTAGTATTGAAAAACTATTTTTCTCAAGGCGATTTTTCGAACTTAGATGTAC
TACGAGATTTTTTTCATGTACTCACTAGGAAGGAGCAAGATGTCCACATTGCTCTTAACATC
AACTTGAAAAGTTTTTTGTGCGAATGGAGACAGAATGCTCTTGTTCTACTAAAAAATACCAA
TAAGTCACACCAAAGGTCTTTTTCAATTCCAAAAGCAACGAAGGATTCTCAAACTCCGTCGG
AAAATAGTGCTGCAACTTTAAAACAAGCTGCAATTGATGCGTATTCTCAAATTCCTGTCATT
CCTTTTTTTCTTCCCAGTGATGTTCCATCTACTAGAACATTTAGTCAGCTTCTAAACACAAAG
CAGGCTTCTTTCCTTGATTTTCTTTCAAAATGTTTGAAATGGGAT

Randomized sequence (SEQ ID NO:17)

ACTAGCCATGGAGACCGAAATGTAAACGTGAAGGAGGCAATATATTATCATGGAGAAATCT
ATTTTAAAAAGAAATGATGAATGAAACTTAACTGAGACATAACATGGTGAGTATCAATTCAAA
AACTATGCGGTATAAAAACATTATGTCAAGCTACATCAGAGGGATCGACGGCACACTTAGA
AGAATTAGTGCTGACATCTCGATGAAACTCGTGTAAAGGGAGCCAGAAGCATGAGAATTGA
TAGTAATACCGCCGACAGTTTTTTATAAAAAATGGAGTTAGTTACCGGACCATATTCATTTAA
ATGTTATGTATAGAATCGCGTCATTGAACCTATTGCAAAAGAGCATCGGATGTTAGTGTCGA
CTGATGAATCTCGGAGTAAATACACTTACGGTTCCTTATAAAAAGGATCATTATTACGGTCT
AATCTATGAAAGTGTTTGCTTAAGAGTTTTGCGTGATTTGACTCCCTGTATCTTGAAAAATAC
AGAGTTTGGTCAAGGATTCTAGTAGTATACTCTTCAATCAATGACTACCTGACCATACTAAA
TCATCGAAATCATCATTTCACTCCAATACTATCAGTATATAACACGCCAGGGACCTCTTACC
CAATTGCCATTAATTGATCAGGCCAAAGGACGTTAATGATACTCCTTTGAATCATAATTGGA
CTCATTGAACAGGTTTTGTTCCTCACCAAAGTTTAATGCGATGATAAATGATATACTCATATA
TTATTAGTTCTGTAATGTATTGCTTGTTCTTTTCGACAACCCCGCAACGT

D) Sequences corresponding to Figure 4. Sequences harbouring synonymous codons relative to the wild type *S. pombe ura4* ORF in the dyad, linker or internal regions of mononucleosomal DNA are indicated in **bold**. The three ORFs have the same size (795 bp) and encode the same Ura4 protein.

ura4 dyad (SEQ ID NO:18)

ATG**GACGCGCGCGTGTTCCAGTCATACAGCGCGCGCGCGGAA**GGGATGAAAAATCCCAT
TGCCAAGGAATTGTTGGCTTTGATGGAAGAAAAGCAAAGCAACTTGTCAGTCGCGGTCGAT
TTGACGAAGAAATCCGAA**ATACTCGAGCTAGTCGACAAGATCGGGCCATACGTATGCGTC**
ATTAAAACGCATATTGACGTTGTCGAGGATTTCGACCAGGATATGGTAGAAAAACTGGTGG
CCTTAGGTAAAAAGCATCGTTTTCTTATCTTTGAGGATCGCAAATTCGCAGACATTGGAAAT
ACC**GTTAAATTGCAGTACGCGAGCGGCGTATATAAGATCGCGAGCTGGGCGCAC**ATCAC
AAATTGCCATACAGTGCCAGGCGAGGGTATTATACAAGGCCTCAAAGAAGTTGGTTTACCT
TTGGGACGTGGTCTCTTGCTT**CTAGCAGAGATGAGCAGTAAGGGAAGCCTAGCGACGGG**
CAGTTATACGGAAAAAACCTTAGAATGGTTTGAGAAGCATACCGATTTTTGCTTTGGCTTTA
TAGCTGGTCGTCGATTTCCTAACCTTCAAAGCGACTACATAACTATGTCCCCTGGTATC**GG**
ACTAGACGTCAAGGGCGATGGATTAGGCCAACAGTACAGGACGCCGGAGGAAGTGATT
GTAAACTGCGGTAGCGATATCATCATTGTTGGTCGTGGAGTCTATGGAGCTGGTCGTAATC
CTGTTGTCGAAGCCAAGAGATATAGAGAAGCTGGTTGGAAGGCATATCAGCAAAGA**TTGA**
GCCAACACTAA

ura4 linker (SEQ ID NO:19)

ATGGATGCTAGAGTATTTCAAAGCTATTCAGCTAGAGCTGAGGGGATGAAAAATCCCATTG
CCAAGGAATTGTTG**GCGCTAATGGAGGAGAAACAGTCAAATCTAAGCGTAGCTGTAGAC**
CTAACGAAGAAATCCGAAATCTTAGAATTGGTAGATAAAATTGGACCCTATGTCTGTGTTAT
CAAGACACATATTGACGTTGTCGAGGATTTCGACCAGGATATGGTA**GAGAAGTTAGTAGC**
ACTCGGCAAGAAACACAGGTTCTTGATATTCGAAGATCGCAAATTCGCAGACATTGGAAA
TACCGTCAAGCTACAATATGCATCTGGTGTGTACAAAATTGCTTCTTGGGCTCATATCACAA
ATTGCCATACA**GTACCGGGTGAAGGCATCATCCAGGGATTAAAGGAGGTCGGCCTGCCG**
TTGGGACGTGGTCTCTTGCTTTTGGCTGAAATGTCTTCCAAAGGCTCTTTGGCTACTGGTT
CCTACACAGAGAAAACCTTAGAATGGTTTGAGAAGCATACCGATTTTTGCTTTGGCTTTAT**C**
GCGGGCAGGAGGTTCCCGAATTTGCAGTCAGATTATATCACGATGAGCCCTGGTATCGG
CTTGGATGTTAAAGGAGACGGGCTGGGACAGCAATATCGTACTCCTGAAGAAGTGATTGTA
AACTGCGGTAGCGATATCATCATTGTTGGTCGTGGAGTCTATGGAG**CCGGCAGGAACCCG**
GTCGTAGAGGCTAAACGCTACCGCGAGGCCGGCTGGAAGGCATATCAGCAAAGACTTTC
TCAGCATTAA

ura4 int (SEQ ID NO:20)

ATGGATGCTAGAGTATTTCAAAGCTATTCAGCTAGAGCTGAGGGGATG**AAGAACCCGATC**
GCTAAAGAGCTATTGGCTTTGATGGAAGAAAAGCAAAGCAACTTGTCAGTCGCGGTCGAT
CTAACTAAAAAGAGTGAGATACTCGAATTGGTAGATAAAATTGGACCCTATGTCTGTGTTA
TCAAGACACATATTGAC**GTCGTAGAAGACTTTGATCAAGAC**ATGGTAGAAAAACTGGTGG
CCTTAGGTAAAAAGCATCGTTTTCTTATCTTTGAGGAT**AGAAAGTTTGCGGATATCGGCAA**
CACCGTCAAGCTACAATATGCATCTGGTGTGTACAAAATTGCTTCTTGGGCT**CACATAACG**
AACTGTCACACCGTACCAGGCGAGGGTATTATACAAGGCCTCAAAGAAGTTGGTTTACCT**C**
TAGGCAGGGGCTTACTATTGCTAGCTGAAATGTCTTCCAAAGGCTCTTTGGCTACTGGTTC
CTACACAGAGAAAACCTTAGAA**TGGTTCGAAAAACACACTGACTTC**TGCTTTGGCTTTATA
GCTGGTCGTCGATTTCCTAACCTTCAAAGCGACTACATAACT**ATGAGTCCGGGCATAGGA**
CTAGACGTTAAAGGAGACGGGCTGGGACAGCAATATCGTACTCCTGAAGAAGTGATTGTA
AAC**TGTGGCTCAGACATTATTATCGTC**GGTCGTGGAGTCTATGGAGCTGGTCGTAATCCT
GTTGTCGAAGCCAAGAGATATAGAGAAGCT**GGCTGGAAAGCGTACCAACAGCGC**CTTTCT
CAGCATTAA

E) Sequences corresponding to Figure 6. Sequence of the synthetic molecules designed to position nucleosomes in *S. pombe* and in *S. cerevisiae*.

S. pombe (SEQ ID NO:21)
CCAGGAAGAATGAGAGCAGGACTCTTCCTCTGAAACCTCGACGGAATTCCAGCAGAAAGT
AAATGTAGTTTAAGAATTAGACCTAGACTTCCTCTCACTTACTCTTGGGCTTGAGGTGAGTG
GGATGCTGGTGGTGATACTCCTGATGATGAACAATCAGCAGCATAAGGAGTCAGAAACAG
CGCATGGGGCACAAACGCAAGCGCAATAGCAAGAGGACCAGCTAAATATTGTAGAAGAGT
TAGATGTGGTCTTCCTCTGCACAATTCGGACAAGACTTTGGGTACTGTTCGTGGTAGGGGT
CCCAGATAAGCACCATGGAAAAGAACAGGAACAAACGCATGAAGAAGCTGACTTTAACATT
TAAGTGTACAAGGAGTAAGACCTTTATATCCTGGTCTTGAAGGCTGAAGGGACATCTGGCA
GACTGCGACTGCTTGTAAGGCTCGTCCTCCTCGTACAAGACTAACAAGAGGACTCAGACA
CATCAGCAGATGCCAGCCACACGGATTATGTGGAACATCAAGAAGAGGATTAGGTTTACCT
GAAGAACAAGTTCTTCGTCTCTCTACTCTTCATGCTCCTGAGGGTGTTCATATTAGAGCTGC
TCCCGGTAGAAGACCAAGAGAAGAACCAGGAAAATGAAGAGCAAGCGTCGCATTCAGCAT
ATAAGTAGTAGTATGGAGATGTAGATTACAATTAATTTTTCATTCTGTCTCTGTTGGTCGTTC
TGGTGCTTTGTCTCATTTTGGTGCTTTTTTAGCTTTTGGACTAGGACCAACATGCTCACACG
CAGTCAGACTCGCAAGAGCACCAGAAGGAGGATTAATGATGAGAAAATTACCTGGTATTGT
TTGTATCTGTGGGAATCGCTTTCTTGAGCGTTTTCGTGATCGTTGTAATTCTGGTTTTTTTTA
C

S. cerevisiae (SEQ ID NO:22)
TATAGTTTCTGACGATACCTAAGAAAAAAACTTACCTAAGCACGCCAAGCCCCCGCGCGAG
GCCCTGGGGGCCCACGGAACGGCATGGGCTGGTGCCAGCCGCCACATCCGTCCAAGGG
TTAGCGGCTTTCCTTTTTTGTATGTACTATAAGATATTTATTCAACGTCTAAGTGATAAAATA
TGTAAAAAAGATGCAACAACGAGCACTCCATCGGGCTGACGCGAGCGGTGCTTTCCAGCA
CACCAGGTTGGTGTGAACCGTCAAGGTGCGACGACCTCCCCGGGCTTTGCCGCATTATAT
GTGTTTAGATAAACGTATTTAAGGATATCTTCTACATGAATATATAGCTTCTAGAAAAAAAAA
CGTGAACGCATCGTGGCGTCGCTACGCAGGGCATCAGTTCGAGGAGCGAGCAGGCCGCG
TAAAGTGCGCCCGTCGCTCCGGGTGCGTCGTTGCGTGTGTTAGATTTTCATTATATGAAAT
TGTTATCTCTCACGATACGCAGGTTATATTAAAAAGCTCAAAGGAGATCCTAAAGCGCCCC
AGGTCCAACCACACAGGAGCACACTCACCAGTCCGAGGGTTTGCGTGGCCGAGCCAACG
GTGGGTGCACGCGGCCACCGTCGTCTGCTTTGATTTTACAACAATTATGAGACATGTTAAT
TAATTTGATAATAAGAATTCTATGAAAGGACGACGCAATGACGCGCCCTACAGCCCACCCT
CCCCCCTCCGATGGGAGGTGGCATGTTGTGAGACGTCGCCCCCTGGCGAGATGTTGAGG
TACTTTGCCTCAGCTATTACATATCTTATTTGTTATTATATTGTGGAAATGACGAGGAAATAA

TGCAAAGCGAAATAAAAAGGCAATACAGACAAGACCGCCACCTCCGCGTTCGAGTACAACT
CCATCGAGTCCACCACACCCTGGCTGTGAGCTGGTCTCGGACGGCTAGGATTTTTTTTTAT
CTTTCTATAATATTAATAACTAGACGTATTAACTCGC

F) Sequences corresponding to Figure 7. Wild-type and modified versions of *S. octosporus and S. japonicus ura4* ORF. In the modified versions, the wild type codons have been replaced by their synonymous codons following the nucleosomal signature of *S. pombe*. The four sequences have the same size (975 bp) and encode the Ura4 protein of *S. octosporus* or *S. japonicus*.

WT *ura4* ORF *S. octosporus* (SEQ ID NO:23)
ATGGATGCTAGACTCAATTACCAATATTCACAAAGAGCAGATGGGTTAAGTAACCCTCTAG
CAAAGGCTTTGCTTGCTTTAATGGAGCAAAAGCAATCGAATTTGTCTGTTGCCGTCGACTT
GACAAAAAAAGCTGACATCCTTGCTCTCATCGAAAAAATTGGTCCTTTTGTTTGTGTCATTA
AGACGCATGTCGATGTCATTGAAGATTTTGATCAAGACTTTGTTCAGCAATTGACCGCCCTT
TCCGAGAAGCATCGATTCTTCATCTTTGAGGACCGTAAGTTTGCCGATATTGGAAACACCG
TCAAGCTCCAATACTCTTCTGGTGTTTACAAAATTGCTTCTTGGGCTCACATTACTAATTGTC
ATACTGTCCCTGGAGAAGGCATTGTTCAGGGTTTGAAAGAAGTTGGTTTACCTCTTGGTCG
CGGCTTGCTTCTCTTGGCGGAAATGTCTTCTAAAGGAAGTCTGGCCACTGGTTCTTACACA
GAAACTACTTTAGAATGGGCTGAAAAACATAGCGACTTTTGCATTGGCTTCATTGCCGGCC
GTCGTTTCCCCAATCTTCAGCATGACTTTATCACTATGTCTCCCGGTATTGGATTGGATGTT
AAAGGCGACGGTATGGGCCAGCAATACCGTACACCTCATGAAGTTATTGTGAATTGTGGCA
GCGATATTATCATCGTTGGCCGTGGAATCTATGGTGCTGGCCGGGATCCCGTAGTTGAAG
CCCAACGTTACAAAAAAGCAGGCTGGGATGCCTACCAAGAACGTCTTACCAAAAAATAA

WT *ura4* ORF *S. japonicus* (SEQ ID NO:24)
ATGTCGGACATCGCCTTGAAAACGTACACGGAGCGCGCCAATGTGCATCCTAATGCAGTC
GCCAAGAAGTTGCTGCGTTTGATGGACGAGAAGAAGTCAAACCTCTCGGTCGCTGTGGAC
TTGACGAAGAAGAACCAAGTCCTGGAACTCGTTGACAAGATTGGTCCCTCGATCTGCTTGT
TGAAGACTCACATTGACATTGTGGAGGACTTTGACGCAGACATGGTCCAACAGCTGGTTGC
TCTGGCTGAGAAGCACAAGTTTTTGATTTTTGAGGATCGCAAGTTTGCCGACATTGGCAAC
ACTGTCAAGCTTCAGTACTCTGCTGGTGTGTACAAGATTGCTTCTTGGGCTGACATCACCA
ACTGCCACACTGTTCCCGGTGAGGGTATTATTAGCGGTTTGAAGGAAGTTGGTCTTCCCTT
GGGTCGTGGTTTGTTGCTTTTGGCTGAGATGTCTTCGAAGGGAACCCTGGCCACTGGAAG
CTACACACAAGCCACATTGGAATTGGCTGAGAAGCACAACGACTTCTGTATGGGTTTTATC
GCCAGACGCCGCTTCCCCGGTCTCAAGAGCGACTTTATTCACATGACACCCGGTGTCGGT
TTGGACGTTAAGGGCGATGGCCTTGGTCAACAATACCGTACACCAGAGGAAGTGATCTGT
GAGAGCCAGAGCGATATCATTATCGTCGGTCGCGGTGTCTACGGCAGCGGCCGTGATGCT
GCTCAAGAAGCTGAGCGCTACAGAAAGGCCGGCTGGGAGGCCTACCAGCGTCGCATTTC
CAAGCAGTAA

Modified *ura4* ORF *S. octosporus* (SEQ ID NO:25)
ATGGATGCTAGATTAAATTATCAATATTCTCAAAGAGCTGATGGCCTCTCGAATCCTCTTGC
TAAGGCTCTTCTTGCTCTTATGGAACAGAAGCAATCAAATCTATCAGTAGCAGTAGACCTCA
CAAAGAAAGCAGACATCCTAGCTCTCATAGAAAAAATAGGACCTTTTGTTTGTGTTATTAAA
ACTCATGTTGATGTGATTGAGGATTTTGATCAGGATTTTGTTCAACAATTGACTGCTCTTTCA
GAAAAACATAGATTCTTCATATTCGAAGACAGAAAATTCGCAGACATAGGAAACACAGTAAA
ACTTCAATATTCTTCTGGAGTTTATAAAATTGCTTCTTGGGCTCACATCACGAATTGTCATAC
TGTTCCTGGTGAGGGTATTGTACAGGGACTAAAAGAAGTAGGATTACCACTAGGAAGAGGA
CTCCTCCTCCTCGCAGAGATGAGCTCAAAAGGATCTTTAGCTACTGGTTCTTATACTGAGA
CTACTCTTGAGTGGGCTGAGAAGCATTCTGATTTTTGTATTGGTTTTATTGCTGGCAGAAGA
TTCCCAAATCTACAACATGACTTCATCACAATGAGCCCAGGAATAGGACTAGATGTAAAAG

GAGATGGAATGGGTCAACAATATAGAACTCCTCATGAGGTGATTGTGAATTGTGGTTCTGA
TATTATTATTGTTGGTCGTGGTATCTATGGAGCAGGAAGAGATCCAGTAGTAGAAGCACAG
AGATACAAGAAAGCAGGCTGGGATGCTTATCAAGAAAGATTAACTAAAAAATAA

Modified *ura4* ORF *S. japonicus* (SEQ ID NO:26)
ATGAGTGATATAGCTTTAAAAACTTATACTGAAAGAGCTAATGTGCATCCGAATGCTGTTGC
TAAGAAACTTCTTCGTCTTATGGATGAGAAGAAATCAAATCTATCAGTAGCAGTAGACCTCA
CAAAGAAGAACCAAGTCCTAGAGCTAGTAGATAAAATAGGACCTTCTATTTGTCTTTTAAAA
ACTCATATTGACATTGTTGAGGATTTTGATGCTGATATGGTTCAACAATTGGTTGCTCTGGC
AGAAAAACATAAATTCCTAATATTCGAAGACAGAAAATTCGCAGACATAGGAAACACAGTAA
AACTTCAATATTCTGCTGGAGTTTATAAAATTGCTTCTTGGGCTGACATCACGAATTGTCAT
ACTGTTCCTGGTGAGGGTATTATAAGCGGACTAAAAGAAGTAGGATTACCACTAGGAAGAG
GACTCCTCCTCCTCGCAGAGATGAGCTCAAAAGGAACTTTAGCTACTGGTTCTTATACTCA
AGCTACTCTTGAGCTTGCTGAGAAGCATAATGATTTTTGTATGGGTTTTATTGCTAGAAGAA
GATTCCCAGGATTAAAATCAGACTTCATCCACATGACGCCAGGAGTAGGACTAGATGTAAA
AGGAGATGGATTAGGTCAACAATATAGAACTCCTGAGGAGGTGATCTGTGAGTCTCAGTCT
GATATTATTATTGTTGGTCGTGGTGTCTATGGATCAGGAAGAGATGCAGCACAAGAAGCAG
AGAGATACAGAAAAGCAGGCTGGGAAGCTTATCAAAGAAGAATTTCTAAACAATAA

G) Sequences corresponding to Figure 8. Wild type and modified sequences of the *kanamycin* ORF. In the modified versions, the wild type codons have been replaced by their synonymous codons following the nucleosomal signature of *S. pombe* or *S. cerevisiae*. The three sequences have the same size (810 bp) and encode the same Kanamycin resistance protein.

Kanamycin ORF (Wild type) (SEQ ID NO:27)
ATGGGTAAGGAAAAGACTCACGTTTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATT
TATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATT
GTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAA
TGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACC
ATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCA
AAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTG
GCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATC
GCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGA
TTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTT
TTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTT
GACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATAC
CAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGC
TTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCG
ATGAGTTTTTCTAA

Kanamycin ORF (*S. pombe* version) (SEQ ID NO:28)
ATGGGAAAAGAAAAAACTCATGTCTCTCGTCCTCGTCTGAACTCGAACATGGATGCTGATC
TTTATGGTTATAAATGGGCTCGTGATAATGTAGGACAATCAGGAGCAACAATATACAGACTC
TACGGAAAGCCAGACGCTCCAGAATTATTTTTAAAACATGGAAAAGGTTCTGTTGCTAATGA
TGTGACTGATGAGATGGTTCGTTTGAATTGGCTTACTGAATTTATGCCTTTACCTACTATTAA
ACATTTCATAAGAACACCAGATGACGCATGGCTCCTCACAACAGCGATCCCAGGAAAAACA
GCTTTTCAAGTTTTAGAAGAATATCCTGATTCTGGAGAGAACATTGTTGATGCTCTTGCTGT
TTTTCTTCGTCGTTTACATTCTATACCTGTTTGTAACTGTCCATTCAATAGTGACAGAGTATT
CAGACTCGCACAAGCGCAGAGCAGAATGAACAATGGATTAGTAGATGCTTCTGATTTTGAT
GATGAGAGAAATGGCTGGCCTGTTGAGCAGGTTTGGAAGGAGATGCATAAATTATTACCTT
TTTCTCCTGATAGTGTAGTAACACATGGAGACTTCAGCCTAGACAACCTCATCTTCGACGAA
GGAAAATTAATAGGATGTATAGATGTTGGAAGAGTTGGAATTGCTGATCGCTATCAGGATC
TTGCGATTCTTTGGAATTGTCTTGGTGAGTTTAGTCCTTCTTTACAAAAAAGATTATTCCAAA

AATATGGAATAGACAACCCAGACATGAACAAGCTTCAATTTCATTTAATGTTAGATGAATTTT
TTTAA

Kanamycin ORF (*S. cerevisiae* version) (SEQ ID NO:29)

ATGGGCAAGGAGAAGACCCACGTGTCCCGCCCCCGCCTGAACTCCAACATGGACGCGGA
CCTTTACGGTTATAAGTGGGCTCGTGATAATGTCGGGCAATCAGGAGCAACAATCTACAGG
CTCTACGGCAAGCCCGACGCGCCGGAGCTGTTCCTGAAGCACGGCAAGGGCAGTGTGGC
CAACGACGTGACCGACGAGATGGTGCGGCTGAACTGGCTGACGGAGTTCATGCCTCTTCC
TACTATTAAACATTTTATCCGTACTCCAGATGATGCATGGCTGCTCACAACCGCCATCCCCG
GCAAGACGGCGTTCCAGGTGCTGGAGGAGTACCCGGACAGTGGGGAGAACATCGTGGAC
GCCCTGGCGGTGTTCCTGCGCCGGCTGCACTCGATTCCTGTTTGTAATTGTCCTTTTAACA
GCGATCGCGTATTTAGATTAGCACAAGCGCAAAGCCGCATGAACAACGGCCTCGTCGACG
CCAGCGACTTCGACGACGAGCGCAACGGGTGGCCGGTGGAGCAGGTGTGGAAGGAGAT
GCACAAGCTGCTGCCGTTTTCTCCTGATTCTGTTGTTACTCATGGTGATTTCTCATTAGATA
ATCTAATCTTCGACGAAGGCAAGCTCATCGGCTGCATCGACGTCGGCAGGGTGGGCATCG
CGGACAGGTACCAGGACCTGGCCATCCTGTGGAACTGCCTGGGGGAGTTCTCGCCGTCG
CTTCAGAAGCGTCTTTTTCAAAAATATGGTATTGATAATCCAGATATGAACAAACTGCAATTC
CACCTCATGCTCGACGAGTTCTTCTGA

Yeast strains and growth conditions

The *S. pombe h- leu 1-32* and *h- leu 1-32 ura4* DS-E (harbouring an internal deletion of the *ura4* ORF) and *S. cerevisiae* aw303-1a (*MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100*) strains were transformed with the appropriate DNA fragments to generate all the mutants used in this work. *S. pombe* was grown at 32 °C in rich medium for the MNase experiments or in minimal medium to select for *ura4* transformants. *S. cerevisiae* was grown at 30 °C in YP medium supplemented with 2% glucose.

Synthesis of modified DNA sequences and yeast transformation

All modified coding and non-coding DNA molecules used in this work were synthesized by GeneArt (Life Technologies) and GeneWiz. *S. pombe ura4* DS-E cells were transformed by electroporation (Forsburg and Rhind, 2006) with modified versions of the *ura4* ORF (Figures 1, 4 and 7) flanked by recombination cassettes. Transformant colonies were directly selected in minimal medium. *S. pombe* 972 h- cells were transformed with DNA fragments, lacking the *ura4* gene, ligated to the *kan* (kanamycin resistance) gene (Figures 2, 3, 6 and 8). Transformants were selected on rich medium plates containing 100 mcg/ml of G-418 antibiotic (except 25 mcg/ml in Figure 8A). A control strain for each mutant strain was constructed by targeting the *kan* gene alone to the same *loci* in *S. pombe* 972 h- cells. *S. cerevisiae* aw303-1a cells were transformed by the lithium acetate protocol and transformants were selected in plates containing 100 mcg/ml of G-418. Correct integration in the targeted *loci* in all transformants was monitored by PCR or by standard DNA sequencing.

Chromatin immunoprecipitation and Q-PCR

ChIP analysis in Figure 9 was performed as described by Pidoux et al. (2004) with some modifications. Exponential *S. pombe* cells were fixed with 1% formaldehyde for 20 min at room temperature. Cells were disrupted in Fast Prep (3 pulses at speed 4.5) and chilled on ice for 2–3 min between each disruption step. Cell extracts were sonicated to shear chromatin to a size range of 200-600 nucleotides using a Diagenode Bioruptor Sonicator (5 cycles of 10 min each with alternating pulses of 30 sec on/off). Samples were incubated overnight at 4 °C with 1 mcg of a monoclonal antibody (8WG16-Abcam) against the CTD repeat (YSPTSPS) of the largest subunit of eukaryotic RNA polymerase II. Samples were purified with the GFX PCR DNA and Gel Band kit (GE Healthcare). Immunoprecipitated chromatin and whole cell extract control samples were resuspended in 70 mcl of sterile water before being used as a template for Q-PCR analysis (95 °C 30 seconds, followed by 40 cycles of 95 °C 5 seconds, 58 °C 30 seconds and 72 °C 15 30 seconds).

Digestion with MNase and indirect end-labelling analyses

400 ml exponential cultures *S. pombe* cells were processed as described by de Castro et al. (2012) except that spheroplasts were resuspended to a final volume of 7 ml in NP-buffer and split into seven fractions of 1 ml. Six samples were digested with 0, 1.5, 3, 7, 15 and 30 units/ml of MNase at 37 °C for 10 minutes. DNA was isolated directly from the remaining sample (naked DNA control) and digested with 0.15 units of MNase under the same conditions. For *S. cerevisiae*, cells from 200 ml exponential cultures were permeabilized as for *S. pombe* excluding the preincubation step, and treated with 10 mg Zymoliase 20T for 10 min at 37 °C. Chromatin was digested with 0, 0.5, 1.5, 3, 4.5 and 6 units/ml of MNase. DNA was purified from the MNase treated samples, digested with the appropriate restriction enzymes, electrophoresed, Southern blotted and hybridized with end terminal probes as indicated in the corresponding Figures.

Preparation of mononucleosomal DNA and next generation sequencing

Mononucleosomal DNA was isolated as described by de Castro et al. (2012) by digestion of chromatin with micrococcal nuclease to generate an 80:20 ratio of mononucleosomes to dinucleosomes, as described by Lantermann et al. (2010). Mononucleosomal DNA was sequenced with an Illumina NextSeq500 platform using the paired-read sequencing protocol. A total of 35.5 million, representing a 354-fold genome coverage (wild-type *S. pombe*), 20.7M, 207-fold (1_6 strain), 24.2M, 242-fold (*shuffled* strain), 29.3M, 293-fold (*S. octosporus*, native), 37.1M, 372-fold (*S. octosporus*, remastered), 28.2M, 283-fold (*S. japonicus*, native) and 22.4M, 224-fold (*S. japonicus*, remastered) reads were aligned to the *S. pombe* genome (ASM294v2.20 assembly 13/08/2013 from PomBase) or to versions of the genome where the wild type *ura4*

ORF had been replaced by the respective *ura4* modified versions (See sequences above) using the NUCwave algorithm (Quintales et al. 2015b).

Generation of the position specific weight matrix

Nucleosomal signatures were defined by the asymmetrical and palindromic distribution of the four nucleotides along 38154, 46120, 27024 and 34526 mononucleosomal DNA sequences of *S. pombe*, *S. octosporus*, *S. japonicus* and *S. cerevisiae*, respectively, aligned to their central (dyad) position (Quintales et al. 2015a). The dimension of the position specific weight matrix (PSWM) is 16 (dinucleotides) x 150 (positions along mononucleosomal DNA). Values for each position (i) and for each dinucleotide (NN) are calculated according to the expression:

$$Score_i^{NN} = log_2 \frac{Freq_i^{NN}}{Freq_{genomic}^{NN}}$$

where $Freq_i^{NN}$ is the frequency of the NN dinucleotide at position i in the group of aligned sequences and $Freq_{genomic}^{NN}$ is the average genomic frequency of the same dinucleotide. Each of the 2384 (16 x 149) elements of the matrix represents the score for each dinucleotide depending on their position along mononucleosomal DNA. Each species generates a different PSWM depending on their different nucleosomal signatures (Figure 5).

Design of sequences for nucleosome positioning on synthetic non-coding DNA molecules

To incorporate the information from *S. pombe* nucleosomal signatures in the non-coding synthetic sequences in Figure 6 we generated six random sequences 153 bp long (147 bp core DNA plus 6 bp linker) (Lantermann et al. 2010) with a 36% genome average G+C content. In the case of *S. cerevisiae* the 6 sequences were 168 bp long (147 bp core DNA plus 21 bp linker) (Lantermann et al. 2010) and with a 38% G+C content. We subjected these individual sequences to reiterate cycles of random single point mutation and selected the resulting sequences after each cycle if they had a higher score in the position specific weight matrix (PSWM) than in the previous cycle. Since unlimited reiteration would generate six identical mononucleosomal sequences, we repeated the process until the average identity between them was not significantly higher than the average identity between individual genomic mononucleosomal sequences (typically 25%). Sequence differences between the six final mononucleosomal sequences in *S. pombe* and *S. cerevisiae* are clearly reflected in their A+T profiles in Figure 6A.

Integration of nucleosomal signatures into ORF sequences by codon substitution

To reproduce the nucleosomal pattern of an endogenous ORF in an unrelated ORF of the same size in *S. pombe*, we identified the midposition of any of the well-positioned endogenous

nucleosomes. From this coordinate, we considered that the midposition of flanking nucleosomes in *S. pombe* would be at a distance equivalent to multiples of 153 bp, as described in the previous section. In a second step, for each codon along of mononucleosomal DNA, we selected the synonymous codon with the highest score in the *S. pombe* PSWM at the corresponding position. The same protocol was applied to *S. cerevisiae*, but in this case, the distance between nucleosomal dyads was 168 bp and the *S. cerevisiae* PSWM was used as a reference. We did not modify wild-type codons corresponding to linker sequences in either yeast. Since neither of the ORFs replaced in *S. pombe* (*ura4* in Figure 7 and *SPBC16G5.03* in Figure 8) and in *S. cerevisiae* (*YKL007W* in Figure 8) encompassed an integer number of nucleosomes, we maintained the ATG and STOP codons and modified only the codons included in the ORF. Wild-type codons in the remaining 150 bp of the two mononucleosomal DNAs including the two ends of the ORFs were not modified.

II. RESULTS

Example 1.- Sequence changes in mononucleosomal DNA destabilize individual nucleosomes

To analyse how robust nucleosomal organization was to changes in the DNA sequence, we selected the genomic region encompassing the *ura4* gene of *S. pombe*. This region is organized in a regular pattern of nucleosomes as shown by partial micrococcal nuclease (MNase) digestion and end-terminal hybridization (Figure 1B, WT). This pattern is consistent with MNase-Seq maps previously described (Soriano et al. 2013, Figure 1A). To modify the sequence of the *ura4* open reading frame (ORF), we replaced the wild-type codons (except for the START and STOP codons) by their synonymous codons such that, when possible, A or T nucleotides were changed to C or G, and vice versa (Materials and methods, sequences corresponding to Figure 1). The resulting ORF (61.5% homologous to the wild type ORF, but encoding the same Ura4 protein) was used to replace the wild-type ORF in its endogenous locus to generate the *S. pombe ura4* 1_6 strain. MNase mapping and hybridization to probe 2 revealed a complete loss of nucleosome positioning where the regular wild-type pattern was replaced by a continuous hybridization smear whose only discernible features were two bands of enhanced sensitivity to MNase in the region occupied by nucleosomes 5 and 6 in the control (Figure 1B, bracket in panel 1_6). This severely altered profile was confirmed by hybridization to probe 1 to visualize the same two regions from the other end of the Hind III fragment (Figure 1C).

To test whether the loss of positioning along the ORF was dependent on the destabilization of the +1 nucleosome, we generated the *S. pombe ura4* 1_3 strain in which only codons

corresponding to nucleosomes 1, 2 and 3 were replaced by the same synonymous codons as in the *ura4* 1_6 strain. In this case, only these three nucleosomes were destabilized, as revealed by the enhanced sensitivity to MNase of the linker regions between nucleosomes 1-2 and 2-3 (Figure 1B, brackets in panel 1_3). We also constructed the *S. pombe ura4* 3_4 strain where synonymous codons were introduced into sequences underlying nucleosomes 3 and 4. In this case, the increased sensitivity to MNase between nucleosomes 1-2 and 2-3 was extended to the linker between nucleosomes 3-4 (Figure 1B, brackets in panel 3_4) while the pattern of upstream or downstream nucleosomes was not detectably affected.

In view of the sensitivity of positioning as regards to codon changes, we wondered whether the distribution of wild-type codons would be relevant to nucleosome positioning. To test this possibility, we generated the *shuffled ura4* strain, where we swapped the different synonymous codons for each amino acid along the wild type *ura4* ORF. The resulting ORF was 77.7% homologous to the wild type, contained the same codons, maintained an identical base composition, and encoded the same protein. However, MNase analysis showed that the regular wild-type nucleosomal array was completely lost in this strain, suggesting that the loss of positioning was due to the modification of the primary DNA sequence and not to changes in the overall base composition of the ORF (Figure 1B, Shuffled). All these mutant strains grew in minimal medium without uracil, and the occupancy of RNA polymerase II along the *ura4* gene was similar in all of them indicating that the observed alterations in the nucleosomal profiles were unlikely to result from major changes in transcription (Figure 9). To confirm these results, we mapped the sites of MNase cutting genome-wide at nucleotide resolution in the *S. pombe* wild type, 1_6 and *shuffled* strains. Results showed that the periodic distribution of MNase sites at linker regions in wild-type cells was replaced by an unstructured and heterogeneous distribution in the 1_6 and *shuffled* strains (Figure 10).

The use of the *ura4* ORF in these experiments raised the question of whether the close link between the DNA sequence and nucleosome positioning would be exclusive to transcribed regions. To test this possibility we selected two groups of two and three well-positioned nucleosomes, in intergenic regions lacking any detectable transcription tested by microarray analysis (Soriano et al. 2013) or by RNA-Seq (Rhind et al. 2011) and substituted their sequences for their randomized versions (therefore maintaining the original base composition) in the same genomic locus (Figure 2A, Materials and methods, sequences corresponding to Figure 2). MNase analysis showed that the positioning of the two groups of nucleosomes was strongly altered relative to the wild-type cells and that this effect was limited to the regions spanning the modified sequences (Figure 2B), as in the case of the *ura4* 1_3 and *ura4* 3_4 strains in Figure 1B. Altogether, these results indicated that modification of mononucleosomal

sequences specifically alter the positioning of nucleosomes associated with them, independently of transcription.

Example 2.- Mononucleosomal DNA encodes portable positioning information

Given the close link between nucleosomes and their underlying sequence, we asked whether mononucleosomal DNA could maintain the positioning of individual nucleosomes in ectopic genomic positions. To address this point, we tried to recapitulate the nucleosomal profile of the *ura4* ORF by assembling individual mononucleosomal sequences from unrelated genomic *loci*. We selected six of these regions from the three chromosomes (sequences and genomic localization in Materials and methods, sequences corresponding to Figure 3) and linked them together to generate a fragment of the same size of the *ura4* ORF (795 bp) (Figure 3A) to replace it in the endogenous *locus*. Since only a part of the sequences underlying nucleosomes 1 and 6 were included in the ORF of the *ura4* gene, only sequences including the equivalent regions relative to the dyad position were selected from two nucleosomes located elsewhere in the genome (Figure 3A). MNase mapping showed that the nucleosomal pattern of the chimaeric construct was virtually identical to the wild-type *ura4* pattern (Figure 3B). Even nucleosomes 1 and 6 were well positioned despite of the chimeric origin of their underlying sequences. As a control, we generated another strain where the sequences of the six ectopic DNA fragments were randomized individually before ligating them together in the same order. In this strain, MNase analysis generated multiple irregularly spaced bands, indicative of the absence of nucleosomal positioning (Figure 3B, random). These results showed that individual mononucleosomal DNA sequences associated with positioned nucleosomes in the genome maintain their positioning potential when transferred to ectopic *loci*.

Example 3.- Positioning information is dispersed across mononucleosomal DNA

The relevance of the DNA sequence in nucleosome positioning raised the question of whether different regions of mononucleosomal DNA would contribute differentially to it. To address this point, we generated three strains where only one third of the mononucleosomal DNA sequence associated with each of the six nucleosomes along the *ura4* ORF was replaced by the same synonymous codons used in the *S. pombe* 1_6 strain (Materials and methods, sequences corresponding to Figure 4). In the *dyad* and *linker* strains, we replaced 51 bp centred on the midposition of mononucleosomal DNA (dyad) or on the linker between adjacent nucleosomes, respectively (Figure 4A, green and orange segments). In the third strain (*int*), we replaced 24-27 bp in the two remaining internal regions of each mononucleosomal DNA between positions -62 to -27 and +27 to +62 relative to the dyad (Figure 4A, black segments). MNase analysis showed that in these three cases internucleosomal bands were slightly more diffused than in the wild type, indicative of a reduction in the affinity between the modified sequences and the histone

octamers (Figure 4B). The similar profile in the three mutant ORFs, suggested that positioning information was not preferentially associated to specific regions of mononucleosomal DNA, but probably depended on the collective contribution of redundant and degenerated elements dispersed along its length.

Example 4.- Engineering nucleosomal positioning on synthetic non-coding DNA sequences

If the DNA sequence plays a significant role in nucleosome positioning, we surmised that it might be possible to design synthetic DNA molecules capable of targeting nucleosomes to specific sites in the genomic context. However, it was not immediately obvious how to design such sequences given the expected degeneracy of the putative sequence determinants. Despite the large variability among the thousands of mononucleosomal sequences in the genome, their aggregated profiles generate well-defined patterns in the distribution of the four nucleotides that we have called nucleosomal signatures (Quintales et al. 2015a). We hypothesized that the information contained within these signatures could contribute to nucleosome positioning and, therefore, we used them as a starting point in the design of the synthetic DNA sequences. To extract the sequence information contained in nucleosomal signatures, we generated a position-specific weight matrix (PSWM), which incorporated the frequency of each of the 16 dinucleotides along the aggregated profiles of thousands of mononucleosomal sequences underlying well-positioned nucleosomes (see Materials and Methods). As a consequence of the species-specific nature of nucleosomal signatures, PSWMs showed different positional values in *Schizosaccharomyces pombe*, *Schizosaccharomyces octosporus*, *Schizosaccharomyces japonicus* and *Saccharomyces cerevisiae* (Figure 5).

Based on this information, we generated six random sequences 153 bp long (147 bp core DNA plus 6 bp linker, Lantermann et al. 2010) and subjected them to reiterate rounds of mutation *in silico* to select those with the highest score relative to the *S. pombe* PSWM but maintaining a sequence identity among them comparable to that between sequences of the same length selected at random in the genome. Sequence differences between the six final mononucleosomal sequences are clearly visible in their A+T profiles as shown in Figure 6A (see Materials and Methods for details on the design of the synthetic sequences).

The resulting 918 bp fragment (Materials and methods, sequences corresponding to Figure 6) was integrated into the intergenic, non-transcribed region between the *S. pombe SPAC6F6.11c* and *SPAC6F6.12* genes (Figure 6B, red arrowhead). MNase analysis across the synthetic fragment revealed a regular array of six nucleosomes, which mapped precisely to the positions predicted by the nucleosomal signatures (Figure 6C, Seq-Sp/Sp). This indicated that

nucleosomal signatures contained information capable of directing nucleosome positioning on the artificial sequences that we generated. Since nucleosomal signatures differ among species, we tested whether the same fragment would also position nucleosomes in *S. cerevisiae*. Insertion into the non-transcribed intergenic region upstream from the *YDL211c* gene in chromosome IV (Figure 6B) followed by MNase analysis showed an irregular banding pattern indicative of the lack of nucleosome positioning (Figure 6C, Seq-Sp/Sc). This result suggested that it might be possible to design sequences based on the *S. cerevisiae* nucleosomal signature capable of positioning nucleosomes in its own genome but not in that of *S. pombe*. To test this possibility, we synthesized a DNA molecule of 1008 bp (Materials and methods, sequences corresponding to Figure 6) based on the *S. cerevisiae* PSWM capable of accommodating six nucleosomes (147 bp core plus 21 bp linker, Lantermann et al. 2010) following the same strategy as used for *S. pombe* (Figure 6A). The resulting fragment was integrated into the same two genomic positions as the previous construct and MNase analysis showed that it generated a perfectly regular array of six nucleosomes at the expected positions in *S. cerevisiae* (Figure 6D, Seq-Sc/Sc), but failed to do so in *S. pombe* (Figure 6D, Seq-Sc/Sp). Altogether, these results show that nucleosomal signatures contain species-specific positioning information capable of targeting nucleosomes to predetermined positions on synthetic artificial DNA sequences.

Example 5.- Nucleosome positioning is not maintained on orthologous sequences of closely related species

Given the incompatibility in positioning between *S. pombe* and *S. cerevisiae*, which are as phylogenetically distant as either of them is from mammals (Sipiczki, 2000), we wondered whether nucleosome positioning would be maintained over orthologous sequences of closely related species. To address this question, we replaced the ORF of the *S. pombe ura4* gene by the orthologous *ura4* ORFs of two species of the same genus, *S. octosporus* and *S. japonicus*. The three ORFs are identical in size (795 bp, Materials and methods, sequences corresponding to Figure 7), have a nucleotide identity of 75.0% and 70.2% relative to *S. pombe*, respectively, and their encoded amino acid sequences are sufficiently similar (82.6% and 74.6% identity) for them to generate functionally interchangeable Ura4 proteins. In addition, the *S. octosporus* and *S. japonicus ura4* ORFs encompass six positioned nucleosomes in their respective genomes at positions comparable to those in *S. pombe* (Figure 7A). Despite these similarities, the sharp internucleosomal bands generated by MNase in the endogenous *S. pombe ura4* ORF (Figure 7B, Native) became more diffuse after its replacement by the *S. octosporus* ORF (Figure 7B, central panel). This was indicative of the increased accessibility of MNase to sequences adjacent to the linker DNA suggesting a less tight interaction between nucleosomes and DNA than in the endogenous *S. pombe ura4* ORF. The effect was more dramatic in the *S. japonicus*

orthologue where multiple dispersed weak bands revealed an extensive loss of nucleosome positioning (Figure 7C, Native). The fact that the *ura4* ORF of *S. octosporus* generated a pattern closer to that of *S. pombe* than *S. japonicus* was consistent with their phylogenetic distance to *S. pombe* of 119 and 221 million years, respectively (Rhind et al. 2011) and with the fact that the nucleosomal signatures of *S. octosporus* and *S. pombe* were more similar than the signatures between *S. japonicus* and *S. pombe* (Figure 5) (Quintales et al. 2015a). These results reinforce the sensitivity of nucleosome positioning to exogenous sequences even in the case of orthologous sequences from species of the same genus.

Example 6.- Engineering nucleosomal positioning on orthologous coding DNA sequences

Given the degenerated nature of nucleosomal signatures we wondered whether they could be incorporated into the coding sequences of the *ura4* ORFs of *S. octosporus* and *S. japonicus,* reconstituting the *S. pombe* nucleosomal pattern. To maintain their native coding specificity, we took advantage of the degeneracy of the genetic code and replaced the codons along the corresponding six mononucleosomal sequences by synonymous codons with the highest possible score at each position in the *S. pombe* PSWM. Despite this restriction in the design of the modified sequences, the resulting ORFs generated a nucleosomal profile indistinguishable from that of the endogenous *ura4* ORF of *S. pombe* (Figures 7B and 7C, Remastered). The nucleotide identity of the modified *S. octosporus* and *S. japonicus* ORF sequences relative to *S. pombe* was 73% and 71%, respectively, which is very close to the 75% and 70% identity of their native versions. This suggests that the overall sequence homology is not a determining factor in the specification of nucleosome positioning. To map more precisely the restoration of the nucleosomal profiles in the modified *ura4* sequences, we generated genome-wide maps of MNase cutting sites at nucleotide resolution of the five strains in Figure 7. Results shown in Figure 12 confirmed that the distribution of MNase cutting sites were slightly more dispersed in the native *S. octosporus ura4* ORF relative to *S. pombe* and that dispersion was much greater along the native *S. japonicus ura4* ORF. Modification of the two ORF sequences (remastered) showed that the distribution of MNase cutting sites was virtually identical to that in the *S. pombe* ORF, coincident with the end-labelling experiments (Figure 7B and 7C).

Example 7.- Engineering nucleosomal positioning on prokaryotic genes

To test whether ORFs completely unrelated to *S. pombe* or *S. cerevisiae* could also be engineered to position nucleosomes at predetermined positions, we selected the prokaryotic *kan* gene, which confers resistance to geneticin. Since no orthologues of this gene are present in the yeast genome, we replaced the endogenous *SPBC16G5.03* and *YKL007W* ORFs of *S. pombe* and *S. cerevisiae*, respectively, by the *kan* ORF. We selected these two ORFs because

they have almost the same size (807 bp) as the *kan* ORF (810 bp), are dispensable for growth in the two yeasts, and have well positioned nucleosomes along their length (Figure 8A). The regular wild type pattern, however, was not maintained after replacing the two ORFs by the native *kan* ORF (Figures 8B and 8C, Kan_WT). To test whether a profile comparable to that of the two endogenous ORFs could be induced over the *kan* ORF, we generated two versions of it by replacing the *kan* native codons by their synonymous codons with the highest score in the *S. pombe* and *S. cerevisiae* PSWM along mononucleosomal DNA (Materials and methods, sequences corresponding to Figure 8). MNase analysis showed that these engineered versions generated regular nucleosomal arrays virtually identical to those of the wild-type *SPBC16G5.03* and *YKL007W* ORFs (Figure 8B, Kan_Sp and Figure 8C, Kan_Sc). To test whether the positioning information was species-specific, we swapped the modified *kan* versions between the two yeasts. Results showed that the regular nucleosome array of the Kan_Sc version in *S. cerevisiae* was completely lost in *S. pombe* (Figure 8B, Kan_Sc) and that the opposite result was obtained when the Kan_Sp ORF replaced the *YKL007W* ORF in *S. cerevisiae* (Figure 8C, Kan_Sp). Altogether, the data shown in Figures 7 and 8 further support the relevance of the DNA sequence on nucleosome positioning. Moreover, they also show that exogenous eukaryotic or prokaryotic ORFs can be engineered in a species-specific manner to direct their packaging into regular nucleosomal arrays indistinguishable from those of the endogenous genes.

III. DISCUSSION

We have focused our work on the contribution of the DNA sequence to nucleosome positioning in the genomic context through two complementary approaches: first, by modifying the sequence in discrete regions and, second, by designing DNA molecules capable of targeting nucleosomes to specific positions.

The first striking finding was the degree of sensitivity of individual nucleosomes to sequence changes, even when the modified regions span only 0.3-1.0 kb (2-6 nucleosomes, approximately). Nucleosome positioning was altered at transcribed (Figure 1) and non-transcribed regions (Figure 2) suggesting that transcription *per se* is not a requirement for nucleosome positioning. This is consistent with the similar RNA pol II occupancy of all the modified versions of the *ura4* gene despite their very different nucleosomal profiles (Figure 1 and Figure 9). The fact that nucleosome positioning has been found to be independent from transcription is in agreement with the previously reported observation that positioning is maintained beyond the transcription termination sites and with the presence of regular nucleosomal arrays in active and inactive versions of many genes during mitosis or meiosis (Soriano et al. 2013).

The precise overlap that has been observed between the length of the modified sequences and the loss of positioning in all the tested cases, suggests that sequence elements contributing to nucleosome positioning could have been disrupted. The existence of such elements was supported by the ability of individual mononucleosomal sequences to direct nucleosome positioning when transferred to ectopic regions in the genome (Figure 3). These sequence determinants, have been suggested to be species-specific based on the observed loss of the regular nucleosomal profile along the *ura4* ORF of *S. pombe* after its replacement by the orthologous ORFs of *S. octosporus* and *S. japonicus* (Figure 7B). Moreover, strong support for the relevance of the sequence to direct nucleosome positioning came from the restoration of the endogenous *S. pombe* pattern after incorporating information from its nucleosomal signature through the use of synonymous codons within the *S. octosporus* and *S. japonicus* ORFs (Figure 7B) and by the generation of regular nucleosomal arrays over prokaryotic genes (Figure 8) and even on synthetic molecules (Figure 6). We conclude from these results that nucleosomal signatures contain positioning information that is only correctly interpreted by the species from which it is derived.

Our findings can contribute to explaining the previous observations that the same DNA sequences are packed differently by nucleosomes of phylogenetically distant species (Bernardi et al. 1992; McManus et al. 1994; Sekinger et al. 2005; Hughes et al. 2012). Furthermore, would be consistent with the recent finding that histones from different organisms have different affinities for the same DNA molecules (Allan et al. 2013). Allan et al. suggesting that different histone octamers might have been adapted through evolution to pack genomes that differ widely in base composition, size, gene density and other structural and functional properties.

Our finding that the swapping of synonymous wild-type codons strongly disrupts the regular nucleosomal profile of the *S. pombe ura4* gene (Figure 1B, shuffled) suggest the intriguing possibility that, in addition to modulating the stability of mRNA and its rate of translation (Plotkin and Kudla, 2011; Presnyak et al. 2015), the distribution of synonymous codons might also contribute to nucleosome positioning along ORFs. This possibility was further reinforced by the restoration of a regular nucleosomal pattern over the *S. octosporus* and *S. japonicus ura4* ORFs by replacing their native codons by synonymous codons closer to the nucleosomal signature of *S. pombe* (Figure 7). A comparable effect was found even on a gene of prokaryotic origin after modifying its wild-type codons according to the nucleosomal signatures of *S. pombe* or *S. cerevisiae* (Figure 8).

Our data also showed, that the information contained in nucleosomal signatures is degenerated and can accommodate a great variety of sequences with similar positioning potential (Figure 6) and is also redundantly distributed along mononucleosomal DNA (Figure 4). The combination of degeneracy and redundancy makes it possible that a great variety of sequences can contribute to nucleosome positioning in the genome.

The potential of nucleosomal signatures to customize nucleosome positioning in coding and non-coding sequences, together with the design of promoters of variable strength, based on their capacity to position or exclude nucleosomes (Curran et al. 2014), opens up the possibility of incorporating this information in the design of synthetic genomes (Annaluru et al. 2014; Haimovich et al. 2015). On a different scale, it will be worth exploring whether the engineering of exogenous sequences to mimic the endogenous nucleosomal pattern of eukaryotic hosts has the potential to improve the expression, maintenance or stability of genes and vectors of biotechnological interest.

References

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC & Pugh BF (2007) Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446:** 572–6

Allan J, Fraser RM, Owen-Hughes T, Docherty K & Singh V (2013) A comparison of in vitro nucleosome positioning mapped with chicken, frog and a variety of yeast core histones. *J. Mol. Biol.* **425:** 4206–22

Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, et al (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science* **344:** 55–8

Badis G, Chan ET, van Bakel H, Pena-Castillo L, Tillo D, Tsui K, Carlson CD, Gossett AJ, Hasinoff MJ, Warren CL, Gebbia M, Talukder S, Yang A, Mnaimneh S, Terterov D, Coburn D, Li Yeo A, Yeo ZX, Clarke ND, Lieb JD, et al (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* **32:** 878–87

Bai L, Charvin G, Siggia ED & Cross FR (2010) Nucleosome-depleted regions in cell-cycle-regulated promoters ensure reliable gene expression in every cell cycle. *Dev. Cell* **18:** 544–55

Beh LY, Müller MM, Muir TW, Kaplan N & Landweber LF (2015) DNA-guided establishment of nucleosome patterns within coding regions of a eukaryotic genome. *Genome Res.* **25:** 1727–38

Berbenetz NM, Nislow C & Brown GW (2010) Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet.* **6:** e1001092

Bernardi F, Zatchej M & Thoma F (1992) Species specific protein--DNA interactions may determine the chromatin units of genes in S.cerevisiae and in S.pombe. *EMBO J.* **11:** 1177–85

de Castro E, Soriano I, Marín L, Serrano R, Quintales L & Antequera F (2012) Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. *EMBO J.* **31:** 124–37

Clapier CR & Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78:** 273–304

Cosma MP, Tanaka T & Nasmyth K (1999) Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97:** 299–311

Curran KA, Crook NC, Karim AS, Gupta A, Wagman AM & Alper HS (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun.* **5:** 4002

Drew HR & Travers AA (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **186:** 773–90

Eaton ML, Galani K, Kang S, Bell SP & MacAlpine DM (2010) Conserved nucleosome positioning defines replication origins. *Genes Dev.* **24:** 748–53

Fath S, Bauer AP, Liss M, Spriestersbach A, Maertens B, Hahn P, Ludwig C, Schäfer F, Graf M, Wagner R.2011. PLoS One. 2011 Mar 3;6(3):e17596.

Forsburg SL & Rhind N (2006) Basic methods for fission yeast. *Yeast* **23:** 173–183

Gaykalova DA, Nagarajavel V, Bondarenko VA, Bartholomew B, Clark DJ & Studitsky VM (2011) A polar barrier to transcription can be circumvented by remodeler-induced nucleosome translocation. *Nucleic Acids Res.* **39:** 3520–8

Gkikopoulos T, Schofield P, Singh V, Pinskaya M, Mellor J, Smolle M, Workman JL, Barton GJ & Owen-Hughes T (2011) A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* **333:** 1758–60

Goodwin S, McPherson JD, McCombie WR (2016) *Nat Rev Genet.* **17**(6):333-51

Gustaffson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M, Protein Expr Purif 2012 ; 83(1):37-46 Haimovich AD, Muir P & Isaacs FJ (2015) Genomes by design. *Nat. Rev. Genet.* **16:** 501–16

Hartley PD & Madhani HD (2009) Mechanisms that specify promoter nucleosome location and identity. *Cell* **137:** 445–58

Hennig BP, Bendrin K, Zhou Y & Fischer T (2012) Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO Rep.* **13:** 997–1003 Hughes AL, Jin Y, Rando OJ & Struhl K (2012) A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol. Cell* **48:** 5–15

Hughes AL, & Rando OJ (2014) Mechanisms underlying nucleosome positioning in vivo. *Annu. Rev. Biophys.* **43:** 41–63

Ioshikhes I, Bolshoy A, Derenshteyn K, Borodovsky M & Trifonov EN (1996) Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.* **262:** 129–39

Kaplan N, Hughes TR, Lieb JD,WidomJ, Segal E (2010) Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. *Genome Biol* **11**:140 - 152

Korber P, Luckenbach T, Blaschke D & Hörz W (2004) Evidence for histone eviction in trans upon induction of the yeast PHO5 promoter. *Mol. Cell. Biol.* **24:** 10965–74

Kornberg RD & Stryer L (1988) Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16:** 6677–90

Koster MJE, Snel B & Timmers HTM (2015) Genesis of chromatin and transcription dynamics in the origin of species. *Cell* **161:** 724–36

Lantermann AB, Straub T, Strålfors A, Yuan G-C, Ekwall K & Korber P (2010) Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat. Struct. Mol. Biol.* **17:** 251–7

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR & Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **39:** 1235–44

Lieleg C, Krietenstein N, Walker M & Korber P. (2015) Nucleosome positioning in yeasts: methods, maps and mechanisms. Chromosoma 124:131-151

Lipford JR & Bell SP (2001) Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol. Cell* **7:** 21–30

Lowary PT & Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276:** 19–42

Mardis E.R. (2008) Next-Generation DNA sequencing methods *Annual Review of Genomics and Human Genetics* **9**: 387-402.

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I & Pugh BF (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18:** 1073–83

Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet.*;**11**(1):31-46. doi: 10.1038/nrg2626. Epub 2009 Dec 8.

McManus J, Perry P, Sumner AT, Wright DM, Thomson EJ, Allshire RC, Hastie ND & Bickmore WA (1994) Unusual chromosome structure of fission yeast DNA in mouse cells. *J. Cell Sci.* **107:** 469–86

Ngo TTM, Zhang Q, Zhou R, Yodh JG & Ha T (2015) Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* **160:** 1135–44

Olson WK & Zhurkin VB (2011) Working the kinks out of nucleosomal DNA. *Curr. Opin. Struct. Biol.* **21:** 348–57

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, Hochwagen A & Keeney S (2011) A hierarchical combination of factors

shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144:** 719–31

Pennings S, Meersseman G & Bradbury EM (1991) Mobility of positioned nucleosomes on 5 S rDNA. *J. Mol. Biol.* **220:** 101–10

Perales R, Zhang L & Bentley D (2011) Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history. *Mol. Cell. Biol.* **31:** 3485–96

Pidoux A, Mellone B & Allshire R (2004) Analysis of chromatin in fission yeast. *Methods* **33:** 252–9

Plotkin JB & Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12:** 32–42

Pointner J, Persson J, Prasad P, Norman-Axelsson U, Strålfors A, Khorosjutina O, Krietenstein N, Svensson JP, Ekwall K & Korber P (2012) CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. *EMBO J.* **31:** 4388–403

Prendergast JGD & Semple CAM (2011) Widespread signatures of recent selection linked to nucleosome positioning in the human lineage. *Genome Res.* **21:** 1777–87

Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR & Coller J (2015) Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160:** 1111–1124

Pugh, B. (2010) A preoccupied position on nucleosomes. *Nature Structural and molecular biology* **17** (8): 923

Quintales L, Soriano I, Vázquez E, Segurado M & Antequera F (2015a) A species-specific nucleosomal signature defines a periodic distribution of amino acids in proteins. *Open Biol.* **5:** 140218

Quintales L, Vázquez E & Antequera F (2015b) Comparative analysis of methods for genome-wide nucleosome cartography. *Brief. Bioinform.* **16:** 576–87

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, Young SK, Furuya K, Guo Y, Pidoux A, Chen HM, Robbertse B, Goldberg JM, Aoki K, Bayne EH, Berlin AM, et al (2011) Comparative functional genomics of the fission yeasts. *Science* **332:** 930–6

Rohs R, West SM, Sosinsky A, Liu P, Mann RS & Honig B (2009) The role of DNA shape in protein-DNA recognition. *Nature* **461:** 1248–53

Sasaki S, Mello CC, Shimada A, Nakatani Y, Hashimoto S-I, Ogawa M, Matsushima K, Gu SG, Kasahara M, Ahsan B, Sasaki A, Saito T, Suzuki Y, Sugano S, Kohara Y, Takeda H, Fire A & Morishita S (2009) Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. *Science* **323:** 401–4

Satchwell SC, Drew HR & Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191:** 659–75

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang J-PZ & Widom J (2006) A genomic code for nucleosome positioning. *Nature* **442:** 772–8

Sekinger EA, Moqtaderi Z & Struhl K (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18:** 735–48

Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.*, **6**(3):e65

Shen CH & Clark DJ (2001) DNA sequence plays a major role in determining nucleosome positions in yeast CUP1 chromatin. *J. Biol. Chem.* **276:** 35209–16

Shim YS, Choi Y, Kang K, Cho K, Oh S, Lee J, Grewal SIS & Lee D (2012) Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. *EMBO J.* **31:** 4375–87

Simpson RT & Stafford DW (1983) Structural features of a phased nucleosome core particle. *Proc. Natl. Acad. Sci. U. S. A.* **80:** 51–5

Sipiczki M (2000) Where does fission yeast sit on the tree of life? *Genome Biol.* **1:** REVIEWS1011

Soriano I, Quintales L & Antequera F (2013) Clustered regulatory elements at nucleosome-depleted regions punctuate a constant nucleosomal landscape in Schizosaccharomyces pombe. *BMC Genomics* **14:** 813

Soriano I, Morafraile EC, Vázquez E, Antequera F & Segurado M (2014) Different nucleosomal architectures at early and late replicating origins in Saccharomyces cerevisiae. *BMC Genomics* **15:** 791

Stockdale C, Flaus A, Ferreira H & Owen-Hughes T (2006) Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J. Biol. Chem.* **281:** 16279–88

Teif VB (2015) Nucleosome positioning: resources and tools online, *Briefings in Bioinformatics*, 1–13Thåström A, Lowary PT, Widlund HR, Cao H, Kubista M & Widom J (1999) Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* **288:** 213–29

Tsankov A, Yanagisawa Y, Rhind N, Regev A & Rando OJ (2011) Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. *Genome Res.* **21:** 1851–62

Warnecke T, Batada NN & Hurst LD (2008) The impact of the nucleosome code on protein-coding sequence evolution in yeast. *PLoS Genet.* **4:** e1000250

Washietl S, Machné R & Goldman N (2008) Evolutionary footprints of nucleosome positions in yeast. *Trends Genet.* **24:** 583–7

Yen K, Vinayachandran V, Batta K, Koerber RT & Pugh BF (2012) Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149:** 1461–73

Yuan DS, Pan X, Ooi SL, Peyser BD, Spencer FA, Irizarry RA & Boeke JD (2005) Improved microarray methods for profiling the Yeast Knockout strain collection. *Nucleic Acids Res.* **33:** e103

Yudkovsky N, Logie C, Hahn S & Peterson CL (1999) Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. *Genes Dev.* **13:** 2369–74

Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS & Struhl K (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.* **16:** 847–52

Zhang L, Ma H & Pugh BF (2011a) Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Res.* **21:** 871-84.

Zhang Z, Wippo CJ, Wal M, Ward E, Korber P & Pugh BF (2011b) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **332:** 977–80

CLAIMS

1. A method for obtaining the DNA sequence pattern determining the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising the following steps:

   a) aligning mononucleosomal DNA sequences of well-positioned nucleosomes of said eukaryotic species, wherein each mononucleosomal DNA sequence is aligned with respect to its central position;

   wherein a mononucleosomal DNA sequence of a well-positioned nucleosome is a DNA sequence which central position does not significantly deviate from its consensus location; and

   wherein the consensus location of the central position of a well-positioned nucleosome is determined by the generation of nucleosome occupancy maps;

   b) obtaining the position-specific weight matrix (PSWM) characteristic for said species, wherein score values for each position (i) and for each k-mer (k) are calculated according to the expression:

$$Score_i^K = log_2 \frac{Freq_i^K}{Freq_{genomic}^K}$$

   wherein $Freq_i^K$ is the frequency of the k-mer at position (i) in the group of aligned sequences and $Freq_{genomic}^K$ is the average genomic frequency of the k-mer;

   wherein the term k-mer refers to an oligomer with a number of base pairs from 1 to n/10, wherein n is the length of the mononucleosomal DNA sequences aligned in step a); and wherein each of the elements of the matrix represents the score for each k-mer in said species depending on its position along the mononucleosomal DNA.

2. The method according to claim 1, wherein the central position of well-positioned nucleosomes in step a) has been determined by the generation of nucleosome occupancy maps by a method comprising the isolation and sequencing of the mononucleosomal DNA.

3. The method according to any of claims 1 or 2, wherein the central position of a well-positioned nucleosome in step a) is defined as peak positions in nucleosome occupancy

maps whose level of occupancy is above the genome average occupancy and the nearest maximum on each direction is at least 120 nucleotides away.

4. The method according to any of claims 1 to 3, wherein the mononucleosomal DNA sequence consists of the 147 bp of the core DNA.

5. The method according to any of claims 1 to 4, wherein in step a) at least 100 mononucleosomal DNA sequences are aligned.

6. The method according to any of claims 1 to 5, wherein said k-mer in step b) has one, two or three nucleotides.

7. The method according to any of claims 1 to 6, wherein said k-mer in step b) is a dinucleotide (NN) and wherein score values for each position (i) and for each dinucleotide (NN) are calculated according to the expression:

$$Score_i^{NN} = log_2 \frac{Freq_i^{NN}}{Freq_{genomic}^{NN}}$$

where $Freq_i^{NN}$ is the frequency of the NN dinucleotide at position (i) in the group of aligned sequences and $Freq_{genomic}^{NN}$ is the average genomic frequency of the same dinucleotide,

wherein each of the elements of the matrix represents the score for each dinucleotide in said species depending on its position along the mononucleosomal DNA.

8. The method according to any of claims 1 to 7, wherein said species is a yeast, preferably wherein said yeast species is selected from the group consisting of *Schizosaccharomyces pombe*, *Schizosaccharomyces octosporus*, *Schizosaccharomyces japonicus* and *Saccharomyces cerevisiae*.

9. A method for tailoring the sequence of a DNA molecule to obtain the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising steps a) and b) according to any of claims 1 to 8, wherein said DNA sequence comprises a coding DNA sequence and said method further comprises:
   i. determining for each codon in the DNA coding sequence the synonymous codon with the highest score in the PSWM of step b); and

ii. replacing each codon in said DNA coding sequence for the synonymous codon determined in (i).

10. A method for tailoring the sequence of a DNA molecule to obtain the nucleosome positioning characteristic of a particular eukaryotic species, said method comprising steps a) and b) according to any of claims 1 to 8, wherein said DNA sequence comprises a non-coding DNA sequence and said method further comprises:

i. identifying the three possible reading frames of trinucleotides in said non-coding DNA sequence;
ii. determining the corresponding codon for each trinucleotide in each of the three possible reading frames;
iii. determining for each codon in each of the three possible reading frames the synonymous codon with the highest score in the PSWM of step b) and replacing it by said synonymous codon; and
iv. selecting from the modified sequences for each of the three possible reading frames obtained in (iii) the one with the highest score in the PSWM.

11. A method for the optimization of a DNA sequence for its expression in a host eukaryotic species wherein said method comprises the tailoring of said DNA sequence to obtain the nucleosome positioning characteristic of said host eukaryotic species by a method according to any of claims 1 to 10.

12. The method according to any of claims 1 to 11, wherein said method is a computer implemented method.

13. A data-processing apparatus comprising means for carrying out the steps of a method of claim 12.

14. A computer program comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of the method of claim 12.

15. A computer-readable storage medium having stored thereon a computer program according to claim 14.

FIGURES

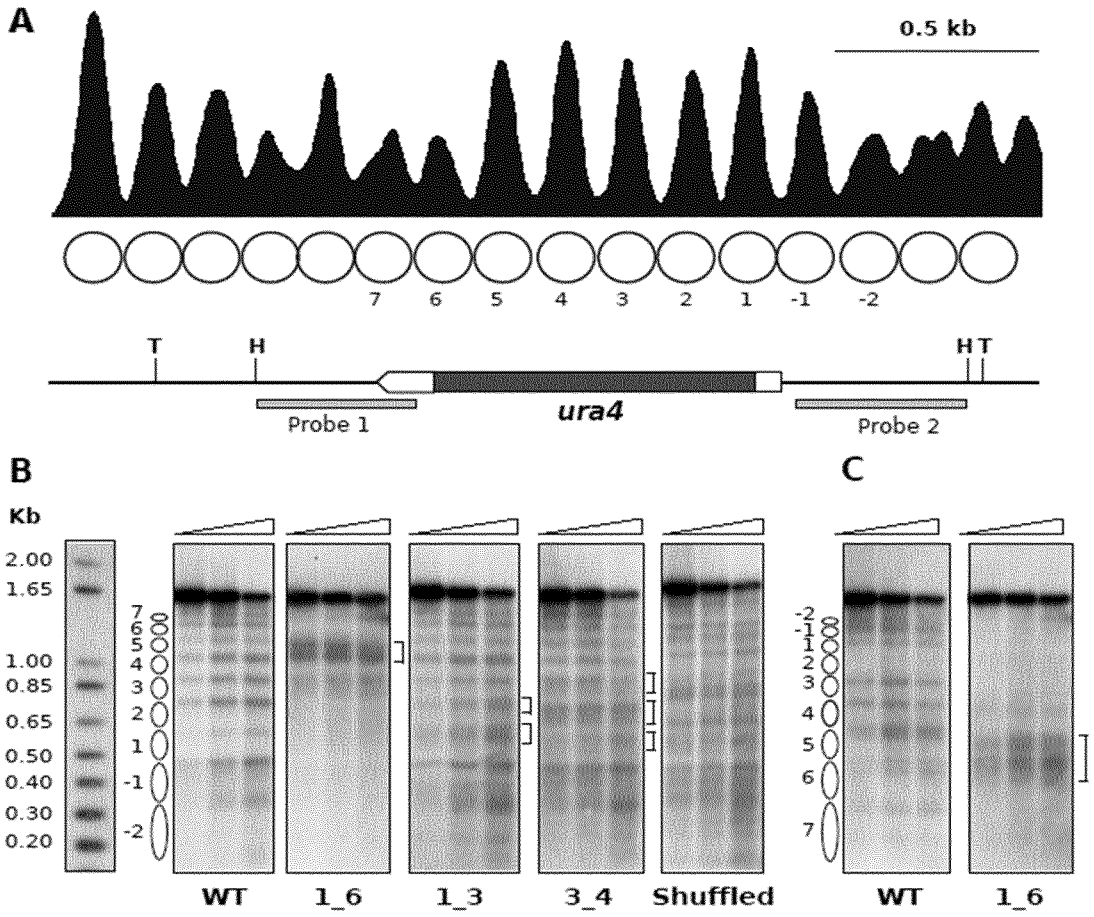

FIG. 1

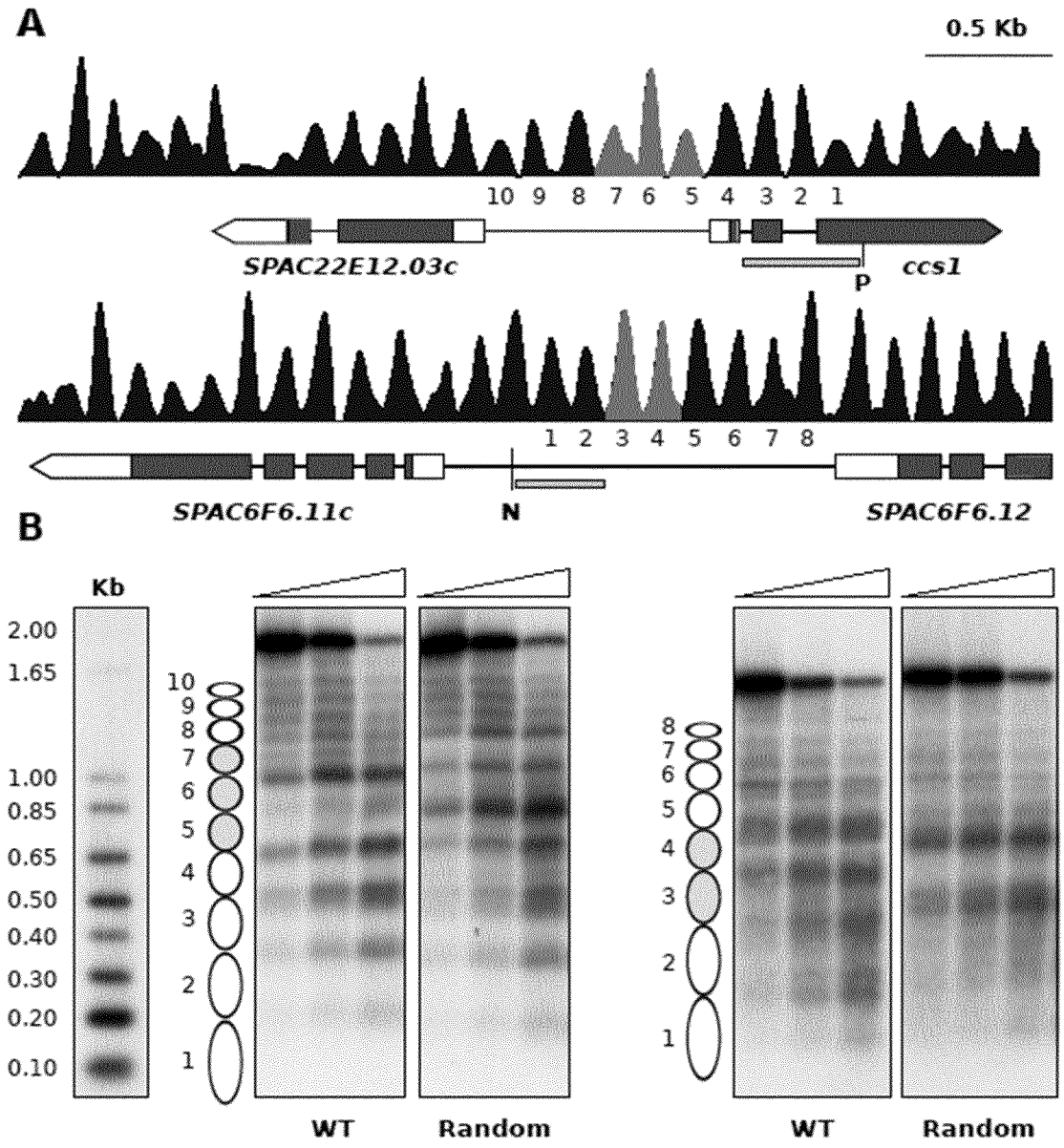
A
0.5 Kb
10 9 8 7 6 5 4 3 2 1
SPAC22E12.03c
P
ccs1
1 2 3 4 5 6 7 8
SPAC6F6.11c
N
SPAC6F6.12
B
Kb
2.00
1.65
1.00
0.85
0.65
0.50
0.40
0.30
0.20
0.10
10
9
8
7
6
5
4
3
2
1
WT
Random
8
7
6
5
4
3
2
1
WT
Random

FIG. 2

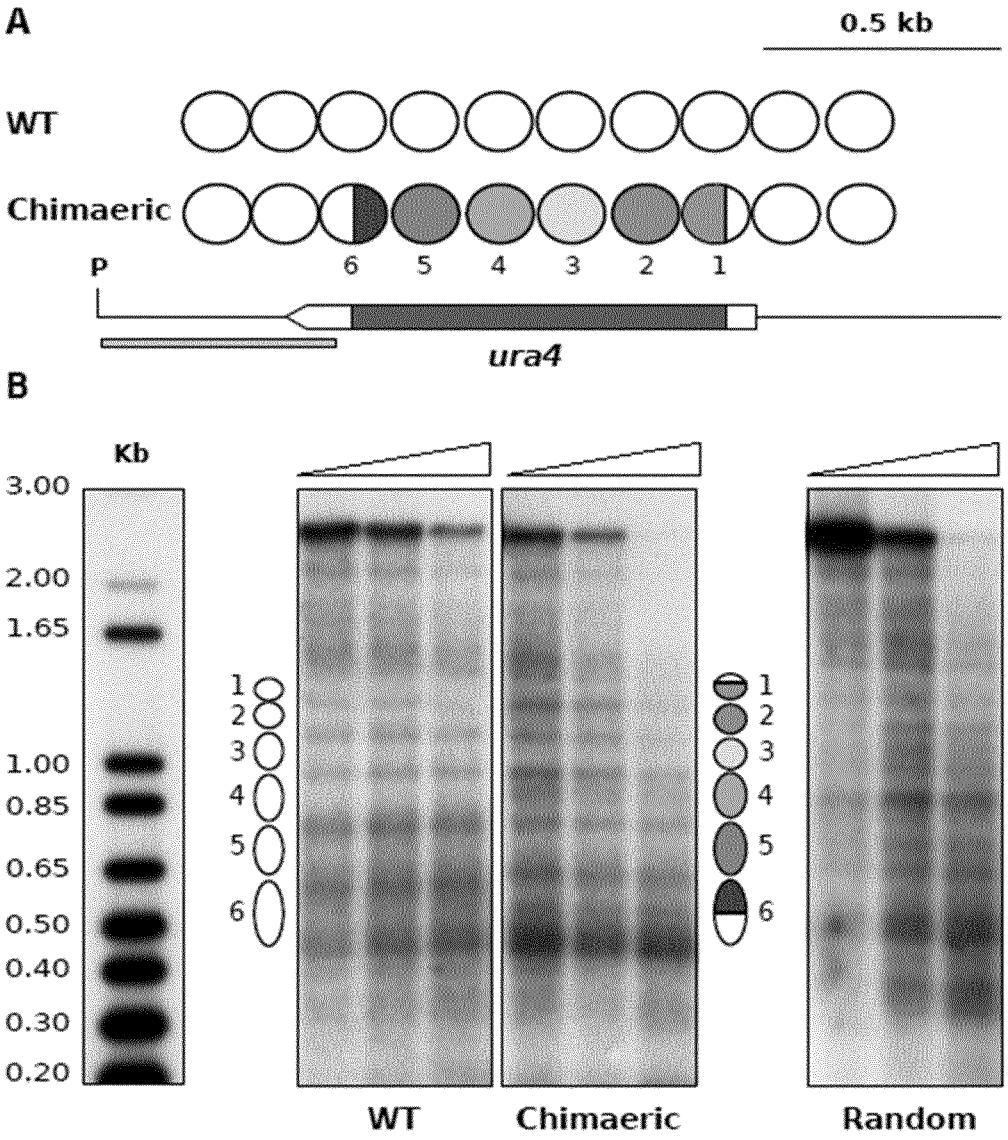
A
0.5 kb
WT
Chimaeric
P
6 5 4 3 2 1
ura4
B
Kb
3.00
2.00
1.65
1.00
0.85
0.65
0.50
0.40
0.30
0.20
1
2
3
4
5
6
WT
Chimaeric
Random

FIG. 3

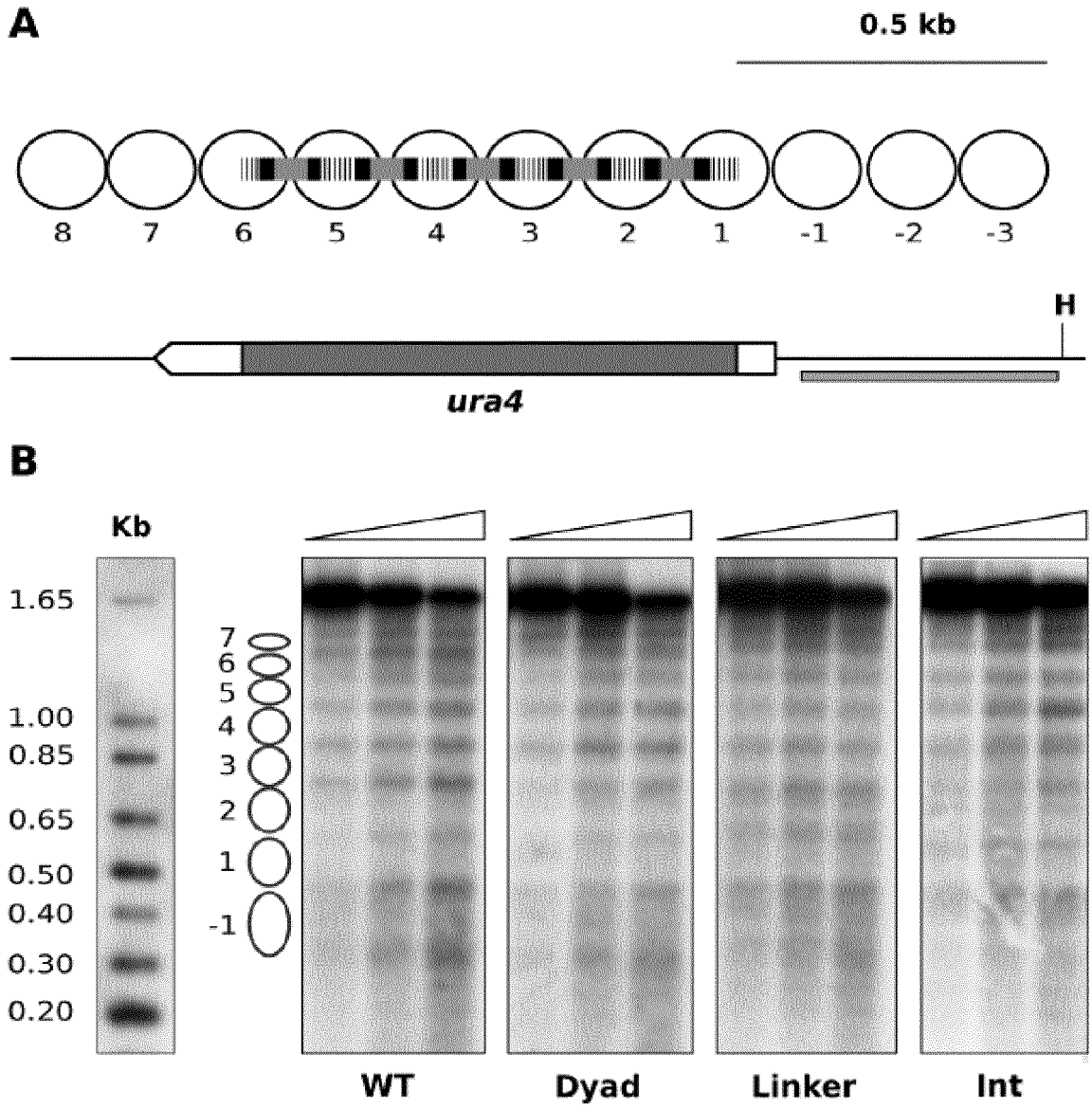
A
0.5 kb
8
7
6
5
4
3
2
1
-1
-2
-3
H
ura4
B
Kb
1.65
1.00
0.85
0.65
0.50
0.40
0.30
0.20
7
6
5
4
3
2
1
-1
WT
Dyad
Linker
Int

FIG. 4

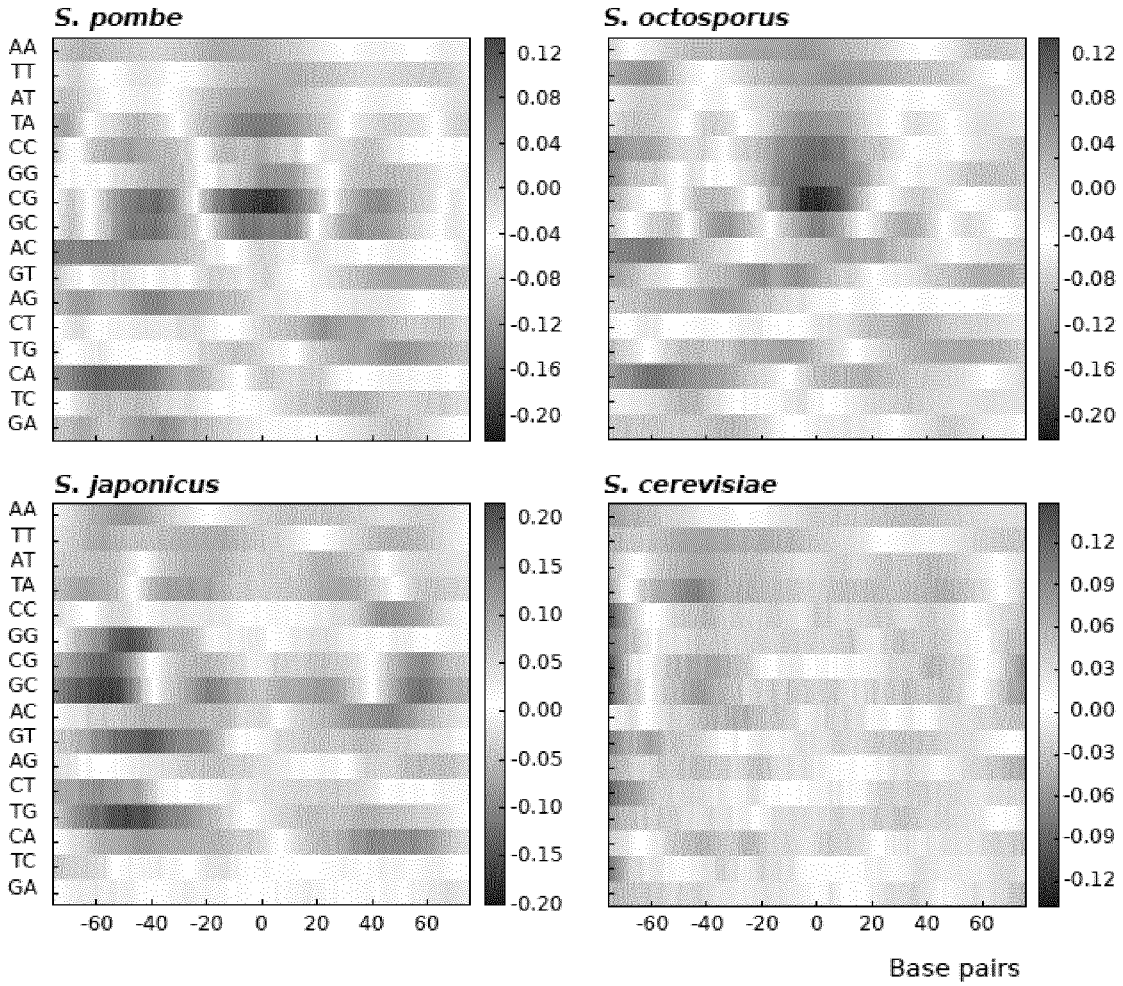
S. pombe
S. octosporus
S. japonicus
S. cerevisiae
AA
TT
AT
TA
CC
GG
CG
GC
AC
GT
AG
CT
TG
CA
TC
GA
0.12
0.08
0.04
0.00
-0.04
-0.08
-0.12
-0.16
-0.20
0.20
0.15
0.10
0.05
-0.05
-0.10
-0.15
0.09
0.06
0.03
-0.03
-0.06
-0.09
-60 -40 -20 0 20 40 60
Base pairs

FIG. 5

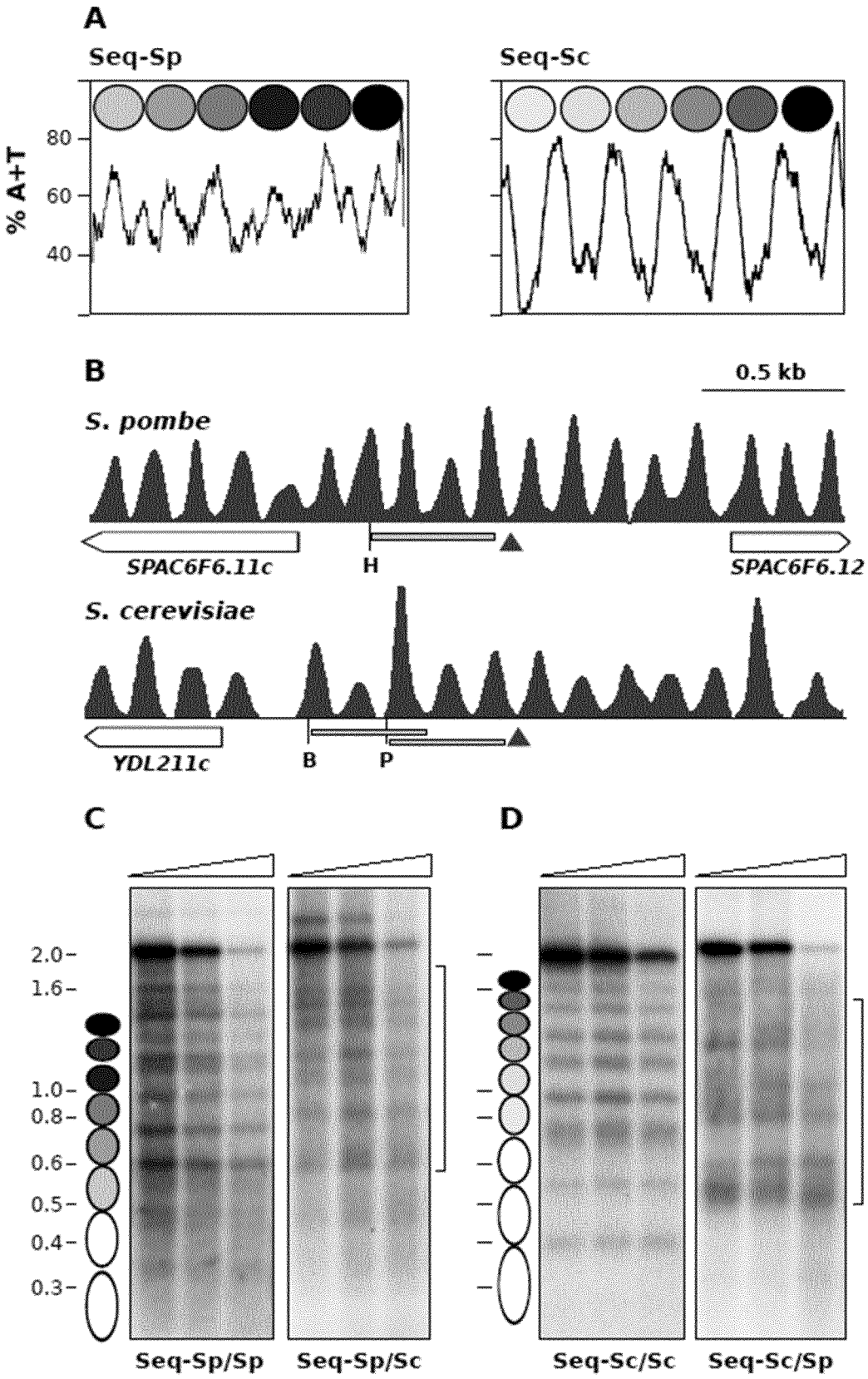
A
Seq-Sp
Seq-Sc
% A+T
80
60
40
B
0.5 kb
S. pombe
SPAC6F6.11c
H
SPAC6F6.12
S. cerevisiae
YDL211c
B
P
C
D
2.0
1.6
1.0
0.8
0.6
0.5
0.4
0.3
Seq-Sp/Sp
Seq-Sp/Sc
Seq-Sc/Sc
Seq-Sc/Sp

FIG. 6

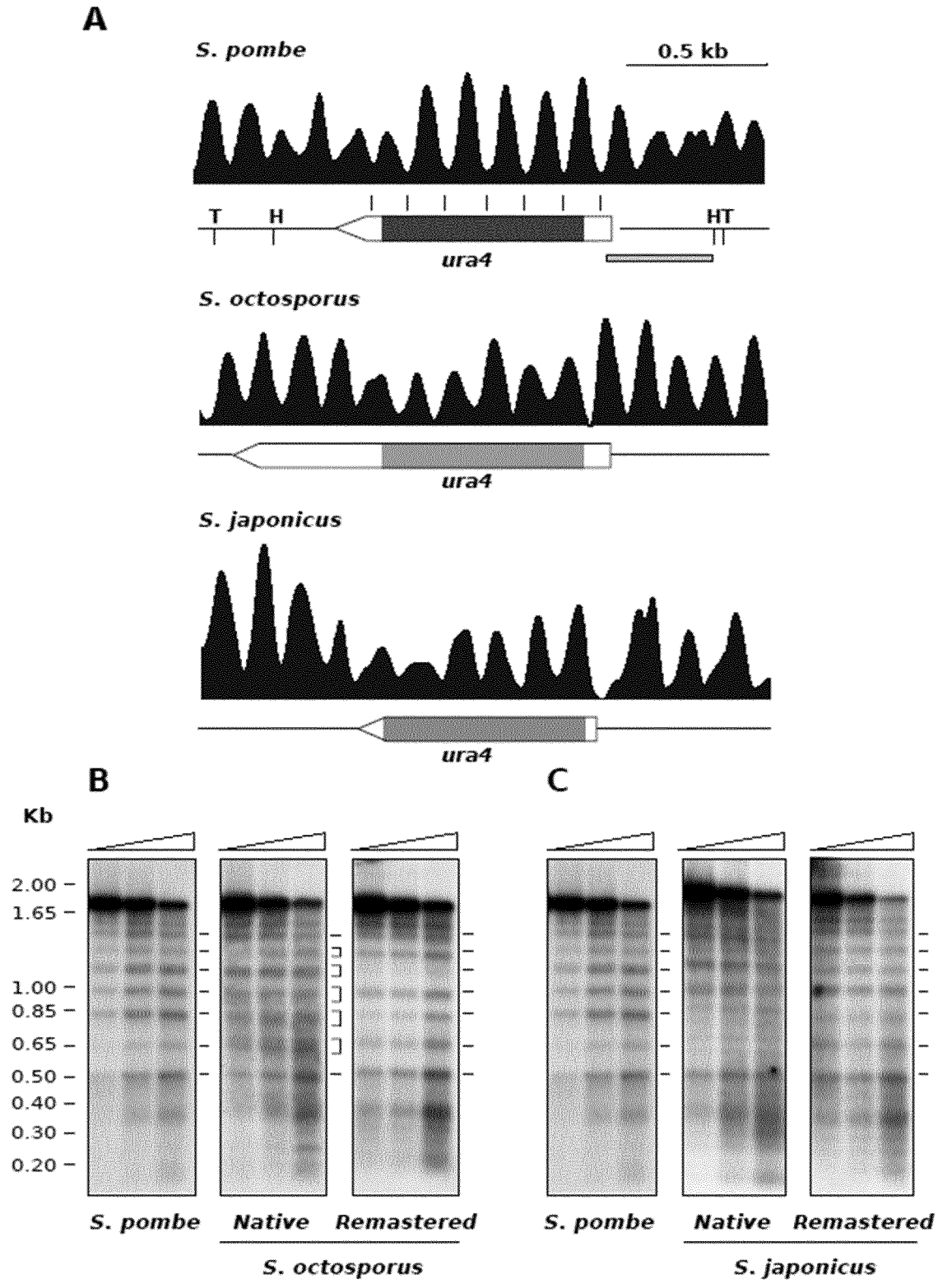
A
S. pombe
0.5 kb
T
H
HT
ura4
S. octosporus
ura4
S. japonicus
ura4
B
C
Kb
2.00
1.65
1.00
0.85
0.65
0.50
0.40
0.30
0.20
S. pombe
Native
Remastered
S. octosporus
S. pombe
Native
Remastered
S. japonicus

FIG. 7

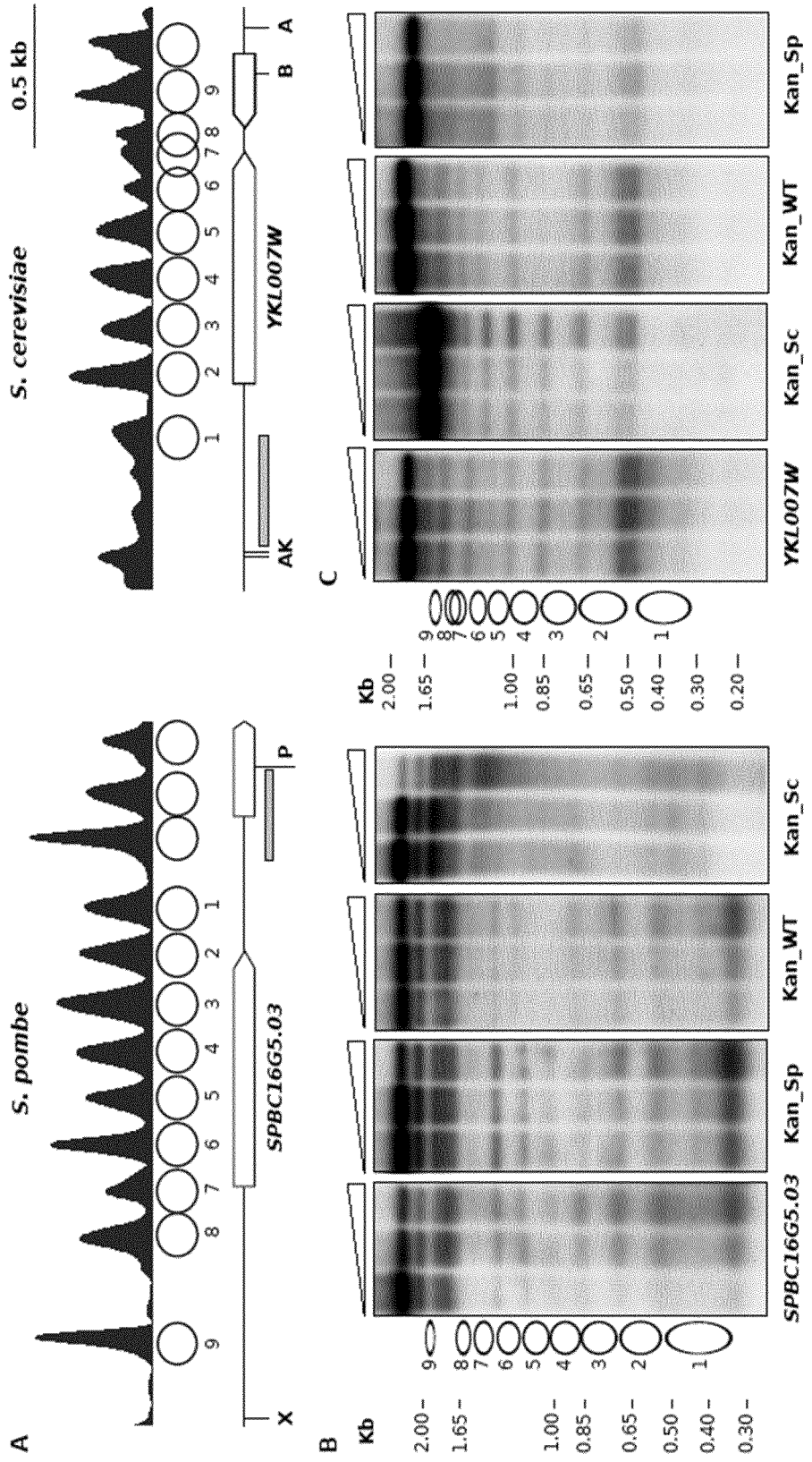
A
S. pombe
9 8 7 6 5 4 3 2 1
SPBC16G5.03
X
P
S. cerevisiae
0.5 kb
1 2 3 4 5 6 7 8 9
YKL007W
AK
B A
B
Kb
2.00
1.65
1.00
0.85
0.65
0.50
0.40
0.30
SPBC16G5.03
Kan_Sp
Kan_WT
Kan_Sc
C
Kb
2.00
1.65
1.00
0.85
0.65
0.50
0.40
0.30
0.20
YKL007W
Kan_Sc
Kan_WT
Kan_Sp

FIG. 8

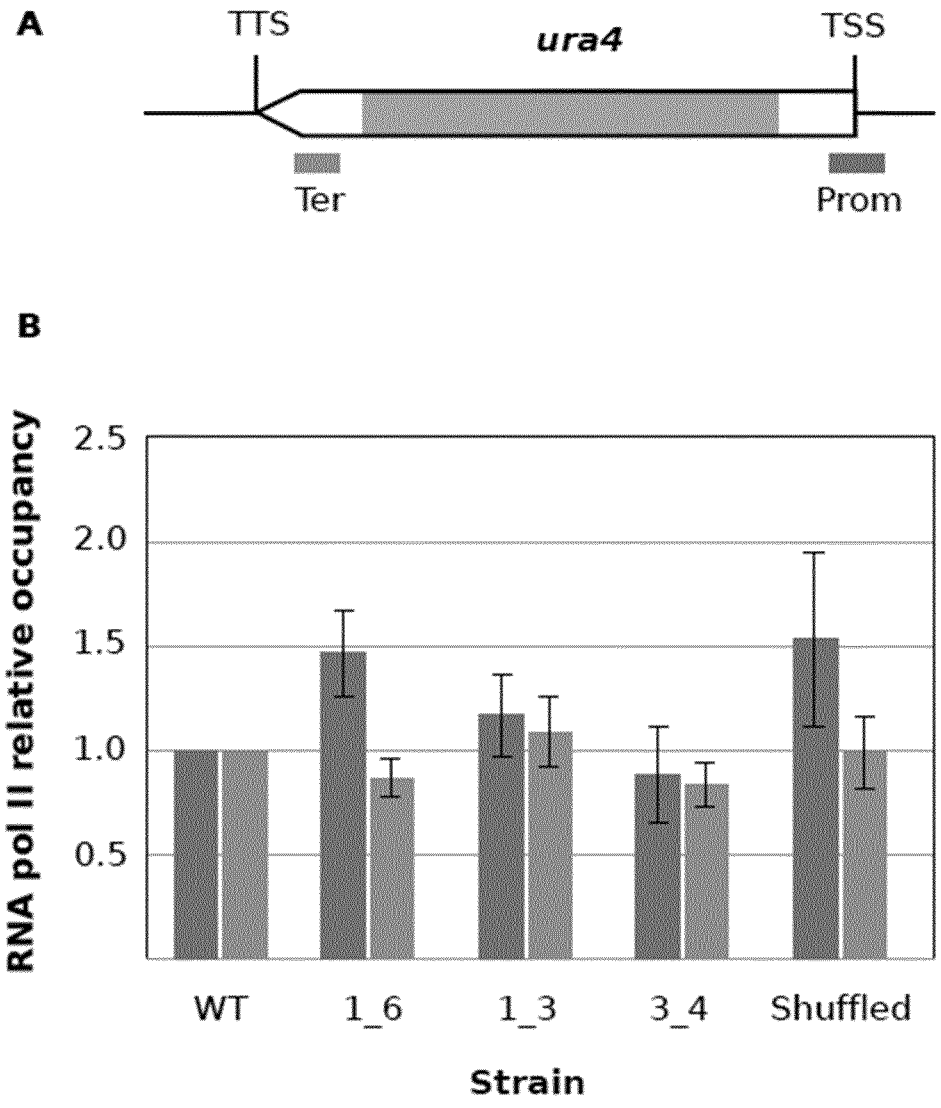
A
TTS
ura4
TSS
Ter
Prom
B
RNA pol II relative occupancy
2.5
2.0
1.5
1.0
0.5
WT
1_6
1_3
3_4
Shuffled
Strain

FIG. 9

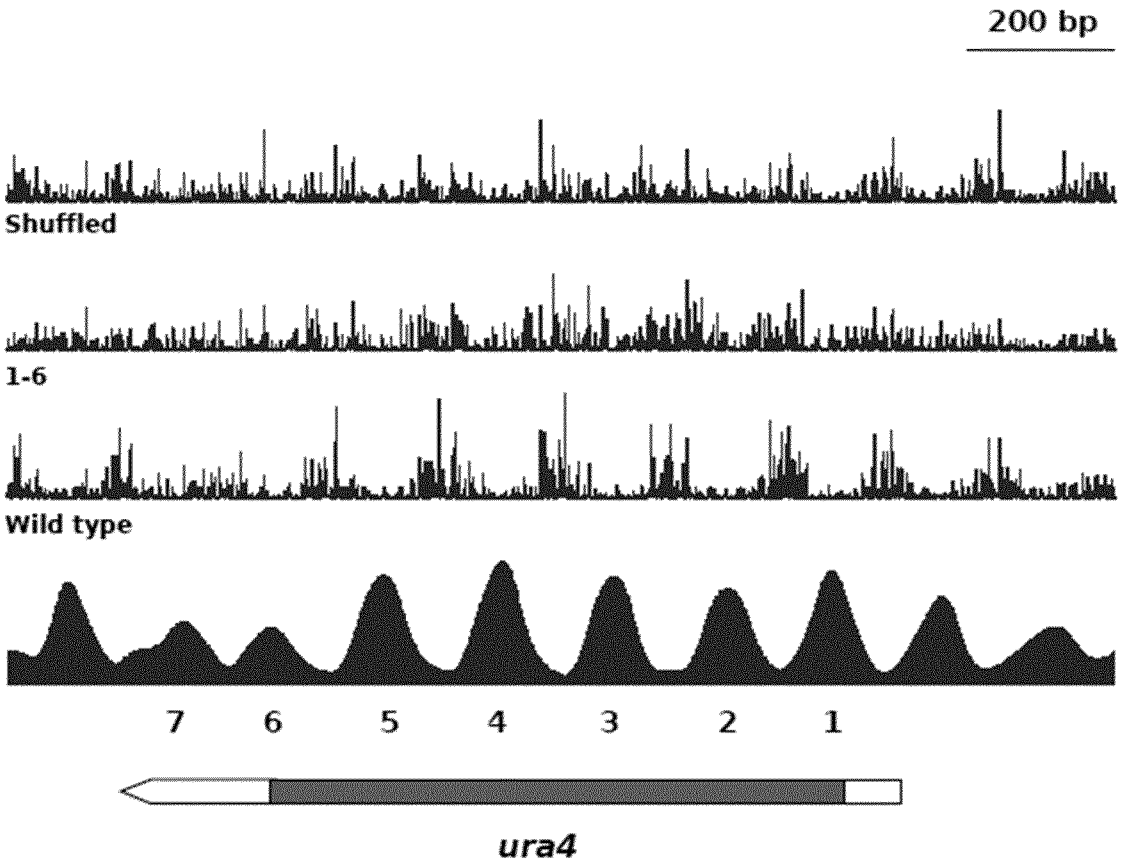

FIG. 10

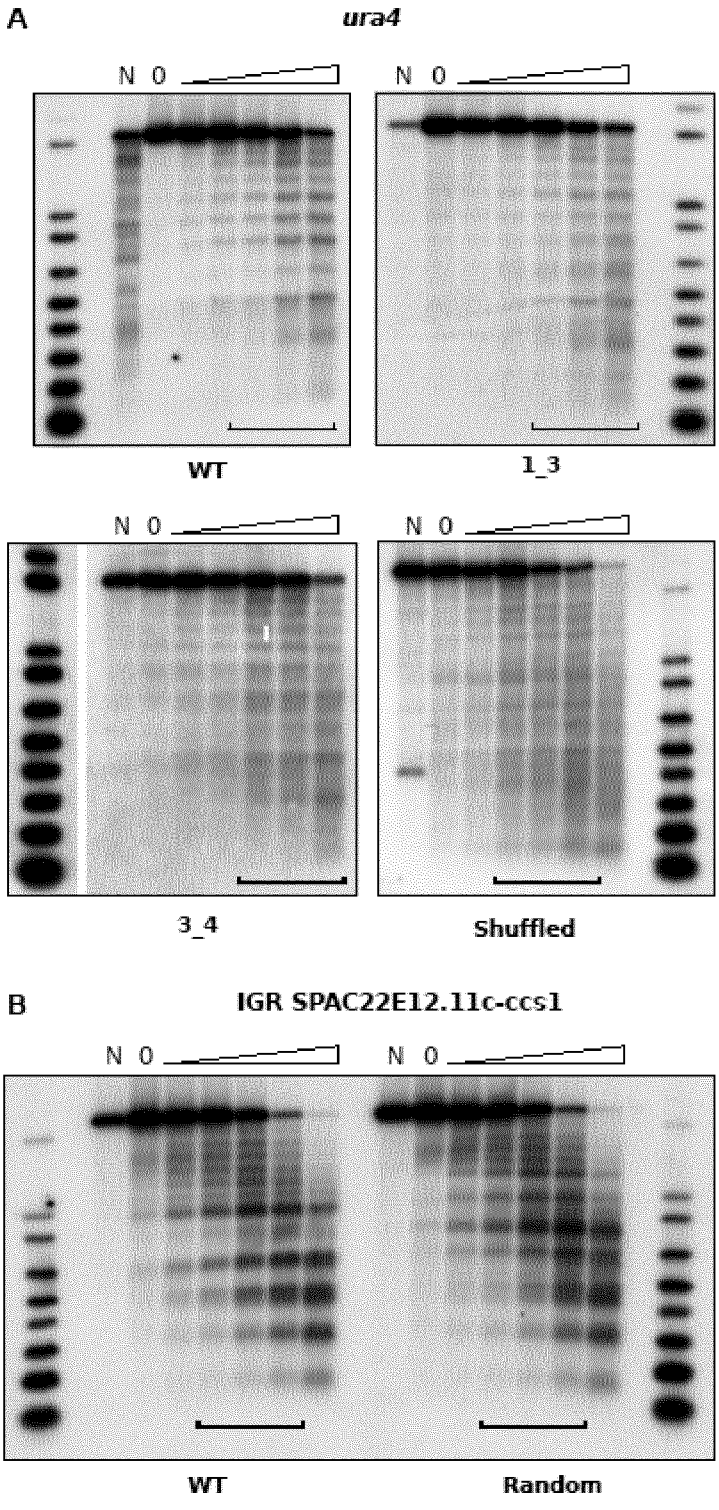
A
ura4
N 0
N 0
WT
1_3
N 0
N 0
3_4
Shuffled
B
IGR SPAC22E12.11c-ccs1
N 0
N 0
WT
Random

FIG. 11

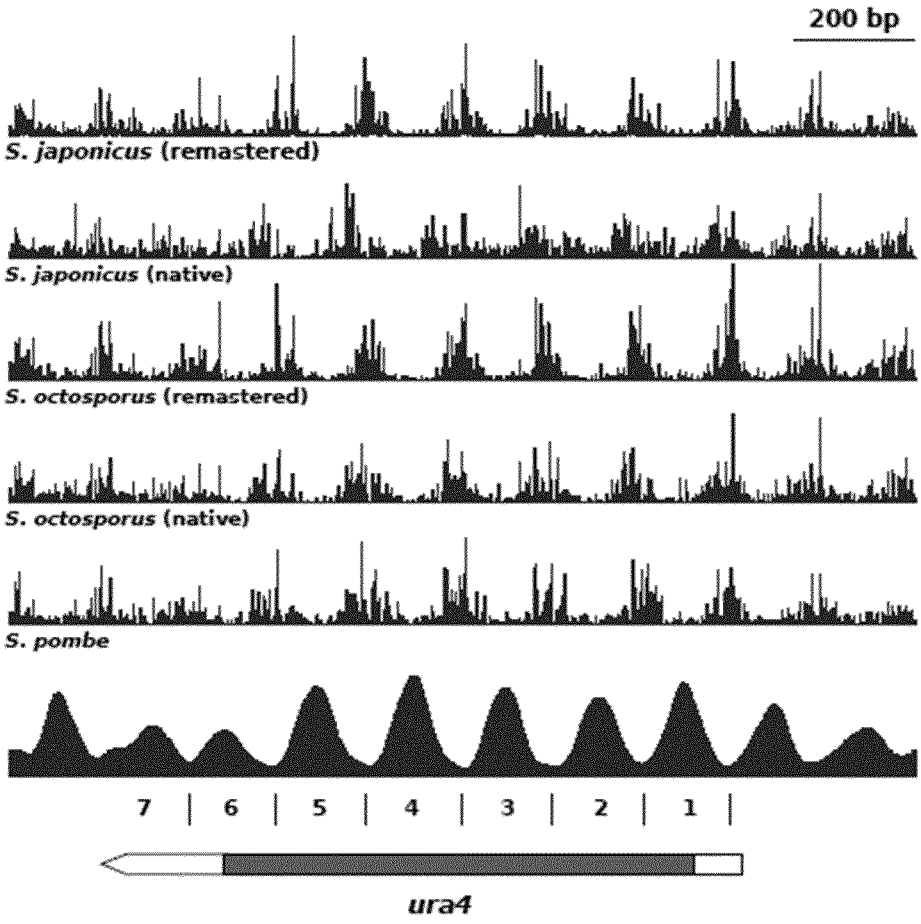
200 bp
S. japonicus (remastered)
S. japonicus (native)
S. octosporus (remastered)
S. octosporus (native)
S. pombe
7
6
5
4
3
2
1
ura4

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/069480

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	L. QUINTALES ET AL: "A species-specific nucleosomal signature defines a periodic distribution of amino acids in proteins", CURRENT OPINION IN STRUCTURAL BIOLOGY, vol. 21, no. 3, 8 April 2015 (2015-04-08), pages 348-140218, XP055333792, GB ISSN: 0959-440X, DOI: 10.1016/j.sbi.2011.03.006 page 2, column 1, paragraph 4 page 2, column 2, paragraph 3 figure 1 abstract the whole document -/--	1-15

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 October 2017	03/11/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Helliot, Bertrand

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	& L Quintales ET AL: "A species-specific nucleosomal signature defines a periodic distribution of amino acids in proteins: Supplementary Figures", Current Opinion in Structural Biology, 8 April 2015 (2015-04-08), XP055333882, Retrieved from the Internet: URL:http://citenpl.internal.epo.org/wf/web/citenpl/citenpl.html [retrieved on 2017-01-10] figure s2 the whole document -----	
T	SARA GONZÁLEZ ET AL: "Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome", GENOME RESEARCH, vol. 26, no. 11, 23 September 2016 (2016-09-23), pages 1532-1543, XP055333794, US ISSN: 1088-9051, DOI: 10.1101/gr.207241.116 page 1541, column 1, paragraph 2 abstract the whole document -----	