



The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis

Juan A. Subirana, Xavier Messeguer

► To cite this version:

Juan A. Subirana, Xavier Messeguer. The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis. *Journal of Theoretical Biology*, 2011, 283 (1), pp.28. <10.1016/j.jtbi.2011.05.025>. <hal-00719494>

HAL Id: hal-00719494

<https://hal.science/hal-00719494v1>

Submitted on 20 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Author's Accepted Manuscript

The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis

Juan A. Subirana, Xavier Messeguer

PII: S0022-5193(11)00267-0
DOI: doi:10.1016/j.jtbi.2011.05.025
Reference: YJTBI6489



www.elsevier.com/locate/jtbi

To appear in: *Journal of Theoretical Biology*

Received date: 29 November 2010
Revised date: 13 March 2011
Accepted date: 17 May 2011

Cite this article as: Juan A. Subirana and Xavier Messeguer, The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2011.05.025](https://doi.org/10.1016/j.jtbi.2011.05.025)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis

Juan A. Subirana^{1*} and Xavier Messeguer²

1. *Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Av. Diagonal 647, E-08028, Barcelona, Spain.*
2. *Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, C. Jordi Girona 1-3, E-08034, Barcelona, Spain.*

Running title

Alternating AT sequences in non-coding DNA and meiosis

* Corresponding author

Tel +34 934016688

Fax +34 934010978

e-mail: Juan.A.Subirana@upc.edu

Abstract

There are general features of chromosome dynamics, such as homologue recognition in early meiosis, which are expected to involve related sequence motifs in non-coding DNA, with a similar distribution in different species. A search for such motifs is presented here. It has been carried out with the CONREPP program. It has been found that short alternating AT sequences (10-20 bases) have a similar distribution in most eukaryotic organisms, with some exceptions related to unique meiotic features. All other microsatellite and repeat sequences vary significantly in different organisms. It is concluded that the unique structural features and uniform distribution of alternating AT sequences indicate that they may facilitate homologous chromosome pairing in the early preleptotene stage of meiosis. They may also play a role in the compaction of DNA in mitotic chromosomes.

Keywords: Non-coding DNA; Microsatellites

1. Introduction

1.1. *Microsatellite frequency in genomes*

Non-coding DNA is very variable in sequence, even when closely related species are compared. Hidden in this variability there are regions which are important for the function of genomes. Comparison of non-coding DNA in different genomes will help to determine if there are some motifs which are conserved in different species, even if an exact sequence conservation is not present. Some of these common sequences might play a general role in chromosome structure and function. It appears from the results presented in this paper that alternating AT motifs may play such a role.

In a previous study it was found [1] that alternating AT sequences (microsatellites) were present in a similar amount in non-coding regions of most eukaryotic genomes. There is no other repeat sequence with this general presence in practically all eukaryotic genomes. Their distribution is analysed in this paper in several species, which differ in the amount of non-coding DNA. The general occurrence of alternating AT motifs, as well as their unique structural properties, suggests that they may play a significant role in chromosome structure. It will be shown that they could be involved in homologous chromosome recognition in meiosis.

Other microsatellite motifs vary much more in different species [1]. Sequences of three or more bases vary enormously in different organisms: they do not appear to be likely candidates for a common role in all species. Among repeats of shorter motifs, C/G and CG are practically absent in most cases. The AC/GT and AG/CT motifs also differ significantly in frequency, over one hundred fold in some cases. Adenine repeats are frequent in most species, with the striking exception of *Trypanosoma brucei*, which has very few of them. However adenine sequences are already known to play a significant role in chromosome structure and in the control of gene expression.

1.2. *Unique structural properties of alternating AT sequences*

Alternating AT sequences are polymorphic. Due to their intrinsic palindromic sequence they may form cruciform structures. At the base pairing level, they may show both

standard Watson Crick and Hoogsteen forms [2, 3]. Mixed structures with both pairing systems have been recently described [4]. Even in standard B form duplexes, they show a unique alternation of low twist TA and high twist AT steps [2]. They also have a narrow minor groove, as it is found in all A/T sequences.

It should be noted that even short alternating AT sequences are absent from DNA signals for specific proteins (restriction enzymes, repressors, transcription factors, regulators, etc.), with a few exceptions to be discussed below. On the other hand, A+T-rich regions are present in matrix attached regions (MARs, reviewed in reference 5) and are the targets for interaction of HMGA proteins [6]. They also appear to be present in the axis of metaphase chromosomes [7]. In all the cases just mentioned, the sequences are not alternating, contain short A-tracts and may involve some C/G bases.

Furthermore, alternating AT sequences have a labile structure: the AT duplex is easily destabilized and unpairs under torsional stress [8, 9].

2. Methods

Sequence data were downloaded from Genbank [10]. They were analyzed with the program CONREPP [11]. It has been briefly discussed elsewhere [12]. The program allows a search of the most frequent short sequences (up to 14 bases) in any whole genome, chromosome or DNA fragment. It also provides a table and a graphical representation of the position of individual short sequences in chromosomes.

The chromosomes used in this study cover a wide biological range, from protozoa and yeasts to higher eukaryotes. All their genomes have been fully sequenced, with the exception of the honey bee, included for its lower CG% among the pluricellular eukaryotes.

3. Results

3.1. Average features in different genomes

The distribution of the (AT)_n motif (with $n \geq 5$) in individual chromosomes of different eukaryotic species is given in Table 1. It has been abbreviated AT5+ throughout the text. Their exact position and statistical distribution for some cases are given in the supplementary data. When two neighbouring AT5+ motifs were found at a short distance, less than 500 bases, it was considered that they formed a single motif. This procedure had a significant effect only in the case of human chromosomes, it increased the average length given in Table 1 by about 50%. The AT5+ motifs appear to be randomly distributed throughout the chromosomes, as it is apparent in Fig. 1. The average frequency of occurrence of AT5+ motifs in all the cases studied is about 100-200 times larger than expected from a statistical point of view. It should be noted that this motif does not appear frequently in exons, in part because it codes for Ile-Phe repeats, which are unusual in proteins. In fact, in all the cases given in Table 1, less than 2% occur in exons. Occasionally long AT repeats are found in some species [1], which may have over 100 bases. However most of the occurrences given in Table 1 are in the range AT5-AT10, usually over 90%. The distribution of repeat frequencies in Table 1 is given as the average length of genome sequence which contains a single motif with at least five consecutive AT units. No correlation with CG% is apparent. Note that in the eukaryotic genomes shown in the table (including *Dictyostelium discoideum*, which has a high percentage of A,T in its genome) this length varies between 8700 and 51000 bases, with some significant exceptions to be discussed below. The numbers in Table 1 will vary significantly if the length of the AT region is changed by a small amount to AT4 or AT6, for example. Different organisms might require different sizes of the alternating AT motif. Also occasional point mutations might be tolerated, in particular if they maintain the RY alternation.

Chromosomes are rather uniformly covered by the AT5+ motifs, although a few gaps are present, as shown in Fig. 1. Coverage is uniform, in the sense that AT5+ motifs are present throughout the genome. However it is not regular, the distance between AT5+ motifs does not follow a standard distribution around the mean values given in Table 1. Instead the distribution is biased towards short distances: it approximately follows an exponential distribution, as shown in detail in the supplementary data. Furthermore there are a few cases in which the distance among neighbour AT5+ motifs is significantly greater than the average value. Most of the large distances are due to the presence of either unsequenced regions or particular structures (centromeres, telomeres, etc.).

An apparently anomalous situation is found in *Apis mellifera*, which presents a higher density of AT5+ motifs. However, if we assume a longer signal in this species, such as AT10, then the density of motifs is 25269 bases, similar to the other cases. It should also be mentioned that a detailed map of microsatellite distribution in honeybees is available [13]. It also shows an uniform distribution, as found by us in other species (Fig. 1 and supplementary data).

3.2. *Caenorhabditis elegans*

This species presents a particular interest and will be analysed in more detail. Besides microsatellite sequences, its genome contains many frequent short repeated sequences [12, 14], but all of them show a low frequency in the centre of the chromosomes. Such repeated sequences appear to be required in order to maintain synapsis in meiotic chromosomes [14]. Terminal poly-guanine sequences may also play a role [15]. In contrast, the distribution of the AT5+ motif is rather uniform throughout the whole chromosomes (shown in Fig. 1). Also the X chromosome, which has very few repeated sequences, has a distribution of AT5+ motifs similar to chromosome I. Thus the distribution of AT5+ motifs is similar to that found in other organisms and differs from the usual terminal distribution of repeats typical in *C. elegans*. However it should be noted that the average distance between neighbouring motifs is appreciably longer than in most of the other species, as shown in Table 1.

3.3. *Exceptions related to unusual meiotic features*

Inspection of Table 1 shows that there are two extreme cases: *Plasmodium falciparum*, with an extremely high density of AT motifs, and *Cyanidioschyzon merolae*, with a very low density. In the latter species half of the chromosomes have no AT5+ motif. These anomalies might be related to their unique chromosomal features. Both species lack chromosomal axes. In both cases mitosis occurs with chromosomes which are not condensed. Furthermore it appears that *C. merolae* only divides in the haploid state, meiosis has not been described. *P. falciparum* usually reproduces in the haploid state, meiosis only occurs at a specific stage of the complex life cycle of this organism. Then meiosis is extremely fast, contrary to what is found in most species [16]. It appears that *P. falciparum* has a unique meiotic mechanism. Also in *P. falciparum* no H1 histone has been detected.

3.4. Correlation with nucleosome positions in yeast

It would be of interest to determine if there is any correlation of the distribution of AT5+ motifs with nucleosome positions. An overall comparison does not appear warranted at this time, since there is a strong controversy on the factors which determine nucleosome positions. However in the case of yeast the results obtained with different programs have been compared and a consensus nucleosome position map is available [17]. Furthermore the yeast genome has a lower proportion of non-coding DNA, when compared with other eukaryotic species. Therefore the rather short chromosome III of *Saccharomyces cerevisiae* has been analyzed in detail. In Table 2 a list of the 23 occurrences of AT5+ found in this chromosome is presented, most of them are found in inter-nucleosomal regions. Only one motif is found in an exon. Their distribution in the chromosome is shown in Fig. 1. It is apparent that AT5+ motifs are distributed throughout the whole chromosome, with a maximum distance between neighbours of about 40 kb (two cases). Although the number of occurrences of the AT5+ repeat is identical to the number of cohesins attached to the chromosome axis [18], their localization does not coincide, as shown in Fig. 2. In fact they show an approximate reciprocal pattern, as expected for AT5+ motifs located in chromosome loops.

Furthermore we have analyzed the composition of the chromosome sequences which appear at both sides of the AT5+ motifs. We find them to be embedded in regions which have a higher proportion of AT sequences than the rest of the chromosome, as it is shown in Table 3. Another feature of the data shown in Table 3 is the presence of frequent AC and GT pairs, next to AT repeats. They maintain the RY alternation of the AT sequence and might also participate in interactions which involve AT sequences.

4. Discussion

4.1 General features

The distribution of AT5+ motifs in eukaryotic chromosomes does not follow any regular pattern. It is rather random, as shown in the supplementary data. However AT5+

motifs show unique features: they are present in practically all eukaryotic chromosomes and cover their whole length. Their average frequency of occurrence is much larger than expected. Such general occurrence and rather uniform distribution in non-coding DNA in different genomes suggests a general structural role for these sequences. Two possibilities will be discussed.

4.2 A role in chromatin packing

Alternating AT sequences may facilitate folding of the chromatin fibres. It has been demonstrated [19] that adequate packing of chromosomes in the nucleus requires kinks with a high flexibility in the chromatin fibres. The unique properties of nucleosome-free alternating AT sequences may provide flexibility. The theoretical estimate of Rosa et al [19] is ten kinks per Mb, which is a lower number than the density of AT5+ sequences reported here in Table 1. It could be argued that somewhat longer AT sequences are required to provide an adequate flexibility. In any case a direct comparison is not warranted, since other flexible nucleosome-free regions may contribute to the flexibility of chromatin fibres.

4.3 Homologue chromosome recognition in meiosis

If there is a general mechanism for the initial recognition process in the preleptotene phase of meiosis in all species, it should be found in non-coding regions. Exons have evolved to optimize protein structure and are not expected to play any specific role in chromosome structure and function. Initial homologue recognition should begin in specific DNA regions found in all organisms and present throughout the whole genome. In Figure 3 is presented a model of two paired sister chromatids in meiotic prophase, based on the description of meiosis given by Kleckner [20]. It is known that interaction between homologous chromosomes takes place at a distance, through accessible regions of the chromatin loops [21-24]. Taken together, the latter observations indicate that in the chromatin loops of meiotic chromosomes there are DNA sequences which favour and stabilize the interaction between homologues. The AT5+ motifs we have described could play such a role. In preleptotene, DNA synthesis has already been finished and sister chromatids remain paired, chromosomes search their homologues [25]. The process usually starts in specific regions which contain repeated sequences, such as

centromeres or telomeres. The sequence of these regions is different from AT5+ and varies in each species. Pairing of homologous telomeres is favoured by their association to the nuclear membrane. In the case of *C. elegans*, pairing centres at the end of chromosomes are present. The proteins and DNA sequences involved have been recently studied by Phillips *et al.* [14]. However, when the usual starting region is damaged, homologous chromosome pairing still occurs [26, 27]. In the study reported here, sequences distributed throughout the whole genome are found. They may facilitate homologous chromosome recognition, even when the starting regions are damaged. Note that in later stages of meiosis (zygotene, pachytene) chromosome pairing is further stabilized by the synaptonemal complex and crossovers are formed. Those later stages have been thoroughly analyzed by Kleckner [20].

The results presented on the distribution of AT sequences suggest that they are likely candidates for homologue recognition in most species. The evidence can be summarized as follows:

- The distribution of AT5+ motifs is similar in most genomes.
- Each AT5+ motif has a different AT-rich sequence around it, as shown in Table 3. This fact should favour adequate pairing, since corresponding regions will not be confused.
- The range of variation in the density of distribution of AT5+ motifs (Table 2) is of the same order of magnitude as it is found in chromosome loops in different organisms [20, appendix]
- In *C. elegans*, the distribution of AT5+ covers the genome in a regular fashion, as shown in Fig. 1, whereas most other repeated sequences are clustered at the ends of chromosomes [12, 24].
- The genome of *C. merolae*, which has no meiosis, lacks AT5+ sequences.
- The genome of *P. falciparum* contains a large number of AT5+, but this organism has an unusually fast meiosis [16].

Direct DNA-DNA interactions might be involved [28, 29]. However such direct interactions require stringent conditions, which might not be available *in vivo* [29]. Alternatively specific proteins may be required in this process, as discussed below. In any case it is worth to point out that DNA with an alternating AT sequences has unique properties, as described in the introduction. Since the interaction of the AT regions is expected to be labile, a trial and error search for exact correspondence would be

facilitated. The fact that AT5+ sequences do not form nucleosomes (Table 2) should also favour their mutual interaction.

The alternative possibility is that specific proteins mediate the interaction between homologous AT5+ motifs. Recognition of such sequences should be expected to occur through the minor groove. There are two families of abundant proteins present in all organisms which interact in such a way with DNA: the TATA box binding proteins (TBPs) and the HMGB family. A single example of each of them has been found in a recent survey of yeast transcription factors [30]. They interact with DNA through a short AT3 region and will be considered in the following discussion.

TBPs have a very specific site of interaction with DNA, which usually includes guanine residues [31]. In yeast, the major TBP, called SpT15, binds to the GNNATATATA motif [30], ahead of the starting codon in genes. In fact, in about half of the sequences shown in Table 3, an SpT15 binding region is present in an adequate position with respect to the starting codon. A much larger number of TBP binding sites are present throughout the whole chromosome. Since these proteins have evolved to play a very specific role in the start of transcription, it is unlikely that they also promote the homologue recognition process. However it can not be excluded that TBPs remain bound to their target DNA sequence during the homologue pairing process. They might also interact with longer alternating AT regions such as those shown in Table 3, in particular since TBPs can form multi-protein complexes.

Members of the HMGB group are the most likely proteins which could participate in homologue recognition. Many of them are sequence specific. They bind to the minor groove of DNA through a cluster of hydrophobic residues [32]. They may cover a long region of DNA through the formation of oligomers [33, 34]. All these properties are consistent with specific binding to alternating AT clusters, in particular if they are unstabilized by torsional stress or cruciform formation: HMGB proteins always bind to distorted DNA regions. It is not possible to point out with certainty which individual protein may play a role in meiosis, since most organisms contain many proteins of the HMGB family, some of them with an unknown function. Furthermore the eventual interaction with long alternating AT sequences has not been studied for any of these proteins. Just as an eventual possibility, it can be mentioned that the Nhp6 proteins in yeast meet all the required properties. They may wrap around DNA duplexes [33] and have a preference for alternating AT sequences [30, Fig.4s]. They are present in the cell in sufficient amount and have multiple roles in chromosome structure and function [35].

They could form complexes with other proteins known to play an essential role in meiotic chromosome pairing in yeast, such as SPO11 [36], MER3 [37], RAD51, the MRX complex [38], etc. However it should be mentioned that these and other known proteins, as reviewed by Hunter [38], only play a role in meiotic recombination after the process of homologue recognition has taken place. Note that no evidence is presently available on the eventual role of any HMGB protein in meiosis.

5. Conclusions

The present study has uncovered a general feature of non-coding DNA: most genomes present a rather uniform distribution of short alternating AT motifs, as shown in Table 1. A few examples are given in Fig. 1. Such uniform distribution suggests a biological significance for these regions. Two possibilities have been considered, which are compatible. Together with other related sequences, they may introduce flexible regions in chromatin fibres and facilitate packing, in particular in mitotic/meiotic chromosomes. They may also play a role in the first stages of homologue chromosome recognition in the preleptotene phase of meiosis. Little is known about proteins which might interact with such alternating AT regions. It is obviously a question which should be studied.

Acknowledgments

This work has been supported in part by grants BFU2009-10380 and TIN2010-21062-C02-01 from the Ministerio de Innovación y Ciencia, Spain. We are thankful to Drs. J.L. Campos and L. Pérez-Grau for suggestions and discussion and to Mr. E. Anokian and Mr. H. M. El-Barody for help in programing.

References

1. Subirana JA, Messeguer X (2008) Structural families of genomic microsatellites. *Gene* 408: 124-132.
2. Abrescia GA, Malinina L, Fernandez LG, Huynh-Dinh T, Neidle S, Subirana JA (1999) Structure of the oligonucleotide d[CGTATATACG] as a site-specific complex with nickel ions. *Nucleic Acids Res* 27: 1593-1599.
3. Abrescia NGA, Thompson A, Huynh-Dinh T, Subirana JA (2002) Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc Natl Acad Sci USA* 99: 1806-2811.
4. Moreno T, Pous J, Subirana JA, Campos JL (2010) Coiled-coil conformation of a pentamidine-DNA complex. *Acta Cryst* 66D: 251-257.
5. Haran TE, Mohanty U (2009) The unique structure of A-tracts and intrinsic DNA bending. *Quart Rev Biophysics* 42: 41-81.
6. Manabe T, Katayama T, Tohyama M (2009) HMGA1 Recognition Candidate DNA Sequences in Humans. *Plos One* 4: e8004.
7. Saitoh Y, Laemmli UK (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell* 26: 609-622.
8. Kohwi-Shigematsu T, Kohwi Y (1990) Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* 29: 9551-9559.
9. Nédjely K, Sýkorová E, Diekmann S, Palecek E (1998) Analysis of a curved DNA constructed from alternating dA_n:dT_n-tracts in linear and supercoiled form by high resolution chemical probing. *Biophys Chem* 73: 205-216.
10. Eukaryotic Genome Sequencing Projects
[<http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>]
11. Algorithmics and Genetics Group [<http://algggen.lsi.upc.edu>]
12. Subirana JA, Messeguer X (2010) The most frequent short sequences in noncoding DNA. *Nucleic Acids Res* 38: 1172-1181.
13. Salignac M, Mougél F, Vautrin D, Monnerot M, Cornuet JM (2007) A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. *Genome Biol* 8: R66.

14. Phillips CM, Meng X, Zhang L, Chretien JH, Urnov FD, Dernburg AF (2009) Identification of chromosome sequence motifs that mediate meiotic pairing and synapsis in *C. elegans*. *Nature Cell Biol* 11: 934-942.
15. Zhao Y, O'Neil NJ, Rose AM (2007) Poly-G/poly-C tracts in the genomes of *Caenorhabditis*. *BMC Genomics* 8:403
16. Sinden RE (1991) Mitosis and Meiosis in malarial parasites. *Acta Leidensia* 60: 19-27.
17. Jiang C, Pugh BF (2009) A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. *Genome Biol* 10: R 109.
18. Blat Y, Kleckner N (1999) Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* 98: 249-259.
19. Rosa A, Becker NB, Everaers R (2010) Looping probabilities in model interphase chromosomes. *Biophys J* 98: 2410-2419.
20. Kleckner N (2006) Chiasma formation: chromatin/axis interplay and the role[s] of the synaptonemal complex. *Chromosoma* 115: 175-194.
21. Rattner JB, Goldsmith MR, Hamkalo BA (1981) Chromosome organization during male meiosis in *Bombyx mori*. *Chromosoma* 82: 341-351.
22. Moens PB (1987) Introduction to Meiosis. In: Moens PB, editor. *Meiosis*. Orlando: Academic Press, pp. 1-17.
23. Scherthan H, Loidl J, Schuster T, Schweizer D (1992) Meiotic chromosome condensation and pairing in *Saccharomyces cerevisiae* studied by chromosome painting. *Chromosoma* 101: 590-595.
24. Weiner BM, Kleckner N (1994) Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. *Cell* 77: 977-991.
25. Zickler D (2006) From early homologue recognition to synaptonemal complex formation. *Chromosoma* 115: 158-174
26. Zickler D, Kleckner N (1999) Meiotic Chromosomes: Integrating structure and function. *Ann. Rev. Genet.* 33: 603-754.
27. Zetka M (2009) Homologue pairing, recombination and segregation in *Caenorhabditis elegans*. In: Benavente R, editor. *Meiosis*. Volff JN. Basel: Genome Dynamics. Karger; pp. 43-55.

28. Danilowicz C, Lee CH, Kim K, Hatch K, Coljee VW, et al. (2009) Single molecule detection of direct, homologous, DNA/DNA pairing. *Proc Natl Acad Sci USA* 106: 19824-19829.
29. Cherstvy AG (2010) DNA-DNA sequence homology recognition: physical mechanisms and open questions. *J. Mol. Recognit.* 24: 283-287.
30. Zhu C, Byers KJRP, McCord RP, Zhenwei S, Berger MF, et al. (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res* 19: 556-566.
31. Patikoglou GA, Kim JI, Sun L, Yang S, Kodadek T, Burley SK (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* 13: 3217-3230.
32. Bewley CA, Gronenborn AM, Clore GM (1998) Minor groove-binding architectural proteins: Structure, function, and DNA recognition. *Annu Rev Biomol Struct* 27: 105-131.
33. Pauli TT, Haykinson MJ, Johnson RC (1993) The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev* 7: 1521-1534.
34. Xiao L, Williams AM, Grove A (2010) The C-Terminal Domain of Yeast High Mobility Group Protein HMO1 Mediates Lateral Protein Accretion and In-Phase DNA Bending. *Biochemistry* 49: 4051-4059.
35. Stillman DJ (2010) Nhp6: A small but powerful effector of chromatin structure in *Saccharomyces cerevisiae*. *Bioch Biophys Acta* 1799: 175-180.
36. Buhler C, Borde V, Lichten M (2007) Mapping Meiotic Single-Strand DNA Reveals a New Landscape of DNA Double-Strand Breaks in *Saccharomyces cerevisiae*. *Plos Biol.* 12: 2797-2808.
37. Storlazzi A, Gargano S, Ruprich-Robert G, Falque M, David M, Kleckner N, Zickler D (2010) Recombination Proteins Mediate Meiotic Spatial Chromosome Organization and Pairing. *Cell* 141: 94-106.
38. Hunter N (2006) Meiotic recombination. In: Aguilera A, Rothstein R, editors. *Molecular Genetics of Recombination*. Heidelberg: Springer-Verlag; pp 381-442.
39. Revenkova E, Jessberger R (2005) Keeping sister chromatids together: cohesins in meiosis. *Reproduction* 130: 783-790.

Table 1 - Distribution of the AT5+ motif in chromosomes

Chromosome	Number of occurrences	Average distance ^a [N _R bases)	N _R bases in chromosome	CG %	Used ^b cases(%)
<i>Homo sapiens 4</i>	12121	23965	187937911	38.2	64.7
<i>Homo sapiens 17</i>	3376	35708	79617833	45.6	65.8
<i>Homo sapiens X</i>	10151	25519	152577922	39.5	58.8
<i>Caenorhabditis elegans I</i>	367	44835	15072418	35.7	91.6
<i>Caenorhabditis elegans X</i>	357	51486	17718851	35.2	96.4
<i>Drosophila melanogaster 2R</i>	1227	19590	21146608	43.3	87.9
<i>Apis mellifera 6^c</i>	3958	5855	17739082	33.9	75.8
<i>Arabidopsis thaliana I</i>	2083	16159	30268597	35.9	90.3
<i>Saccharomyces cerevisiae III</i>	23	13766	316617	38.5	100
<i>Saccharomyces cerevisiae IV</i>	69	22202	1531918	37.9	100
<i>Saccharomyces cerevisiae (all)</i>	561	23213	12070821	38.3	92.7
<i>Schizosaccharomyces pombe (all)</i>	375	34632	12534385	36.1	96.5
<i>Trypanosoma brucei (all)</i>	3438	9008	25789175	46.5	82.5
<i>Dictyostelium discoideum (all)</i>	4460	8746	33929333	22.4	86.6
<i>Plasmodium falciparum (all)</i>	51164	1689	22859932	19.1	18.0
<i>Cyanidioschyzon merolae (all)</i>	17	973338	16546745	55.0	100

a. Average distance between neighbouring AT5+ motifs. Each motif is an exact sequence of five or more AT doublets. Closely spaced motifs are counted as a single motif.

b. Cases used (%) in order to calculate the average distance. All distances smaller than 500 bases have been excluded, as discussed in the text.

c. This chromosome has an estimated length of 17.7 Mb, but only part of it has been sequenced.

Table 2 - Position of AT5+ motifs in Chromosome III of *Saccharomyces cerevisiae*

AT _N	9204 n	AT _N	91288 N	AT _N x2	247979
AT _N	12334	AT _N	108895	AT _N	259304
AT _N	22386 N	AT _N	153972	AT _N	260024
AT _N	46972	AT _N	190271	AT _N	273407
AT _N x2	50321	AT _N	199492 G	AT _N x2	292570
AT _N	52402 N	AT _N	213867	AT _N	293124 n
AT _N x2	62749	AT _N	224294	AT _N x2	315516 N
AT _N	65872	AT _N x2	224771		

AT_N indicates a position with N=5-9. When there are two contiguous motifs with N=5 the factor x2 has been added.

N indicates that the sequence is found in a nucleosome with 100% probability, as reported in the visualization browser of Jiang and Pugh [17].

n indicates that the sequence is found in a nucleosome with a lower probability [17].

G indicates that the sequence is found in an exon.

The neighbouring sequences for all these sites are given in Table 3.

Table 3 - Neighbouring sequences of AT5+ motifs present in *S. cerevisiae*, chromosome III.

Pos. 9217:	tttttttgggttaa ATATATATATAT ac ATATATAT ag ATAT ag ATAT acac ATgtATagATgtAT
Pos. 12341:	AT gttt ATgtAT ttttgcgaa ATATATAT ATATATAT AT tctacacag ATATAT ac ATAT ttgttt
Pos. 22385:	tctaagtaaacgaa ATgtgtgtgtttct ATATATAT AT agac ATATAT gaagg AT tcctagcggt
Pos. 46971:	ga AT gtaggt AT tagtt AT ta AT ggaggtg ATATATAT ATAT gtt AT tATATAT ttgcATATAT a AT a
Pos. 50336:	gcacgcgtaacaaa ATATATATATATATAT ATATATAT ATgtATgtccAT acgggttttcgtttgt
Pos. 52409:	tacttacc AT cttctctcttt ATATATAT ATATATAT ATgtATAT tttcagtg ATAT ac AT ac AT t
Pos. 60787:*	aaacac AT gcttg AT ac ATAT ac ATAT ac ATATATAT aagctc AT t AT aaacctgtgtgc AT ggaaa
Pos. 62748:	ttgaagaagt AT gtacgt AT gc AT acacac ATATATAT ATATATATATAT agtac AT tt AT ctctt
Pos. 65871:	ctct AT tc AT tcaagactgt AT ttgtttg ATATATAT ATgtATATAT ac ATAT t AT tttcgtagtg
Pos. 91287:	ctaacttttctactttt AT ttcagca ATATATAT ATATATAT ttcaagg ATAT acc AT tcta AT
Pos. 94503:*	gcgggta AT agggtgc AT a AT agt AT tgta ATATATAT aa ATATAT aaagctact ATgAT ag ATAT tt
Pos. 108894:	taa AT aagttc ATAT t AT a AT cttttagc ATATATAT ATATATAT tg AT ctctctctcttt AT tt
Pos. 153971:	c AT gtctct AT tttttgcacaaa ATAT ac ATATATAT ATAT aagtaagtg ATAT tactactgaga
Pos. 190270:	AT taccttacgtgttagtgact ATAT t ATATATAT ATATAT gt ATATATAT aaagggaggagttt
Pos. 199499:	AT gttt ATgtAT ttttgcgaa ATATATAT ATATATAT tctacacag ATATAT ac ATAT ttgttttt
Pos. 213890:	aggtct ATATATATAT tct ATAT ac ATAT tt ATATATAT t AT tagaactttaca ATAT agt ATAT acca
Pos. 224293:	ac ATAT ttct ATATATAT gt ATATATAT ct ATATATAT ATATAT teccagcgt ATAT gactgcacaaga
Pos. 224778:	cttcta AT tt AT ctttaaa AT t ATATATAT ATATATAT ATATAT gtgtgtttgt AT tactctgtggga
Pos. 247986:	ATATAT tt ATgtATAT tt ATgtATATATAT ATATATAT ATATAT gcgta AT t AT gcag AT ttctcctt
Pos. 254246:*	ttttttcttctttt AT tagt ATgtAT ac ATATATAT tactttt ATAT ttcttt AT gcgc AT ttta
Pos. 259303:	gcggct AT ggac AT ttttaaac ATATAT tt ATATATAT ATacATATAT gcgt ATAT tg AT aa AT cac
Pos. 260031:	taa ATAT aagaaaaag AT agt ATATATAT ATATATAT ATacATAT ac ATAT aaaaggaaagtaagct
Pos. 273406:	aaaa AT tc AT aaaa AT aaggacca ATAT aa ATATATAT AT tt AT gtgctagtgcac AT cctagt AT
Pos. 292569:	ttttca AT a AT a AT ttt AT aa AT tagt AT ca ATATATAT ATATATATATATAT tt AT ttgtttactttt
Pos. 293131:	AT gttt ATgtAT ttttgcgaa ATATATAT ATATATAT AT tctacacag ATATAT ac ATAT ttgtttt
Pos. 315523:	AT aac AT t ATAT c ATAT ttct ATATATAT ATATATAT ATATATATATATAT c ATAT aacgggtgaaa AT

The regions are centred around an AT4 motif, shown between vertical slashes. The position of the first base of the latter motif in the chromosome is given. The thirty bases before and after each AT4 motif are shown. AT pairs are enhanced. Some sequences which do not contain the AT5+ motif (Table 2), but have neighbouring AT regions have an *

Figure captions

Fig. 1. Distribution of AT5+ motifs in four different chromosomes. Note the different scale for each species. The drawing has been obtained with the CONREPP program [12].

Fig.2. Distribution of AT5+ motifs (black rectangles) and cohesion binding sites (red dots) in yeast chromosome III, as determined by Blat and Kleckner [18]. Protein coding regions are shown as green arrows. The exact position of AT5+ motifs is given in Table 2.

Fig. 3. Homologous chromosome pairing in the early stages of meiosis. A simplified model of the structure of sister chromatids in meiotic prophase is shown in the upper scheme. A single chromosome axis or axial core is usually observed, formed by the association of the two chromatids (reviewed by Moens [22]). The axial core is indicated in the figure by the two parallel lines in the centre. Two regions may be distinguished in the axis: **a**, where the two ends of each chromatin loop are associated and **b**, the chromatin stretch between consecutive loops. It is not clear if the **b** regions are mainly composed of protein or contain long stretches of MARs (Matrix Attached Regions). In the loops we indicate by asterisks (*) the regions which are involved in homolog recognition. The number of such regions in each loop is expected to vary. In this paper the DNA sequences which might be present in the latter regions are investigated. Sister chromatids are maintained together, probably by cohesins (reviewed by Revenkova and Jessberger [39]). They are indicated as black squares in the axial core. In the cell the chromatin loops are not extended, they are covered by histones and condensed. In the lower part of the figure, a scheme of the interaction between the two homologues is shown. The recognition sequences indicated with asterisks are in contact. In this simplified model only one sister of each chromosome participates in the interaction.

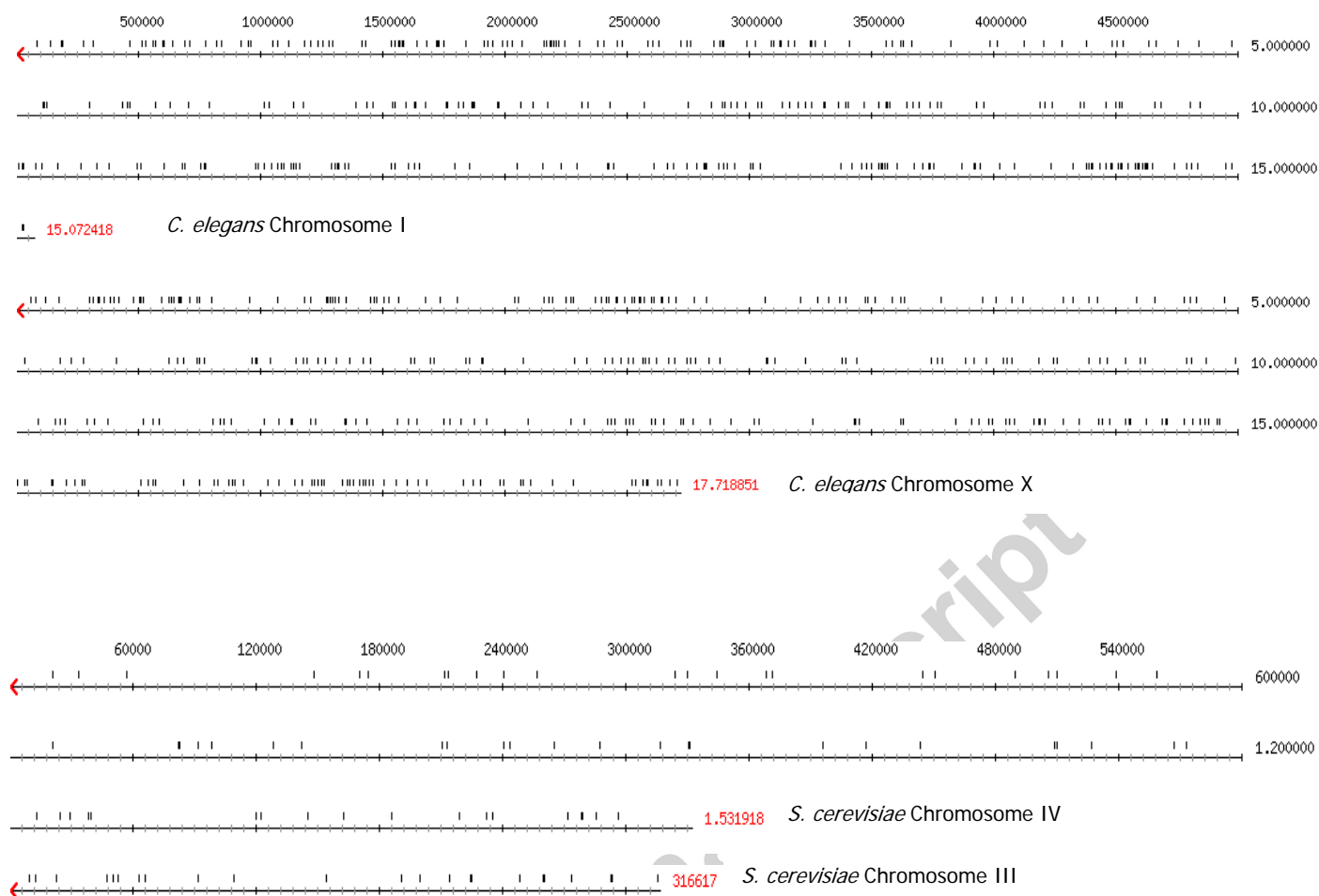


Figure 1

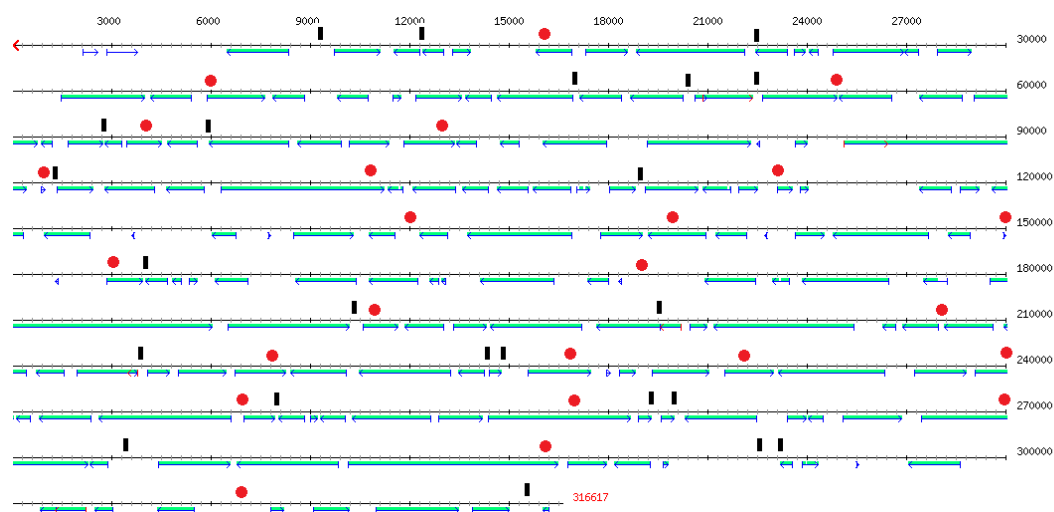


Figure 2

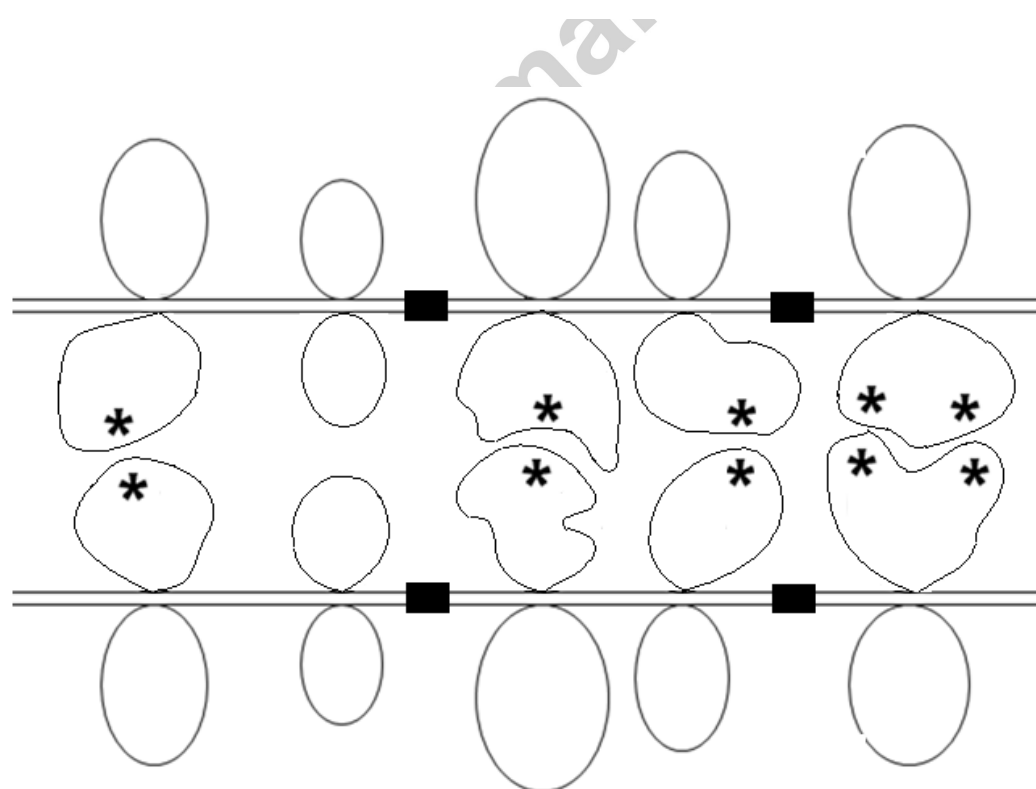
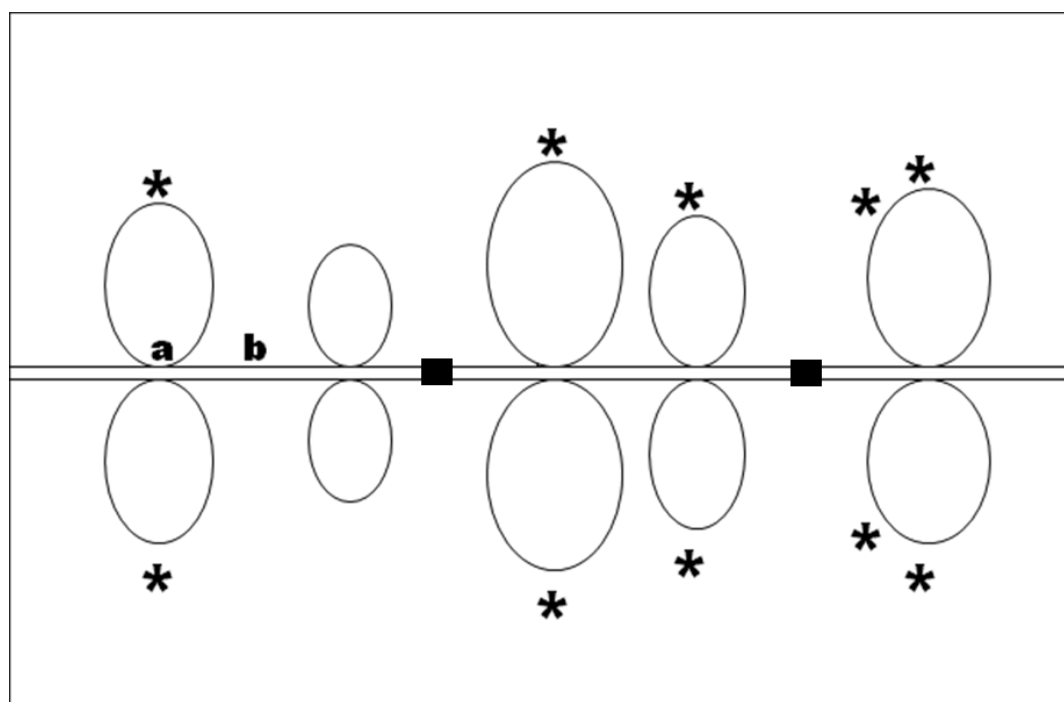


Figure 3

- We analyze the distribution of alternating AT microsatellites in several genomes. > We find a similar distribution of alternating AT sequences in most eukaryotic species. > Alternating AT sequences may facilitate homologous chromosome pairing in meiosis. > Proteins which recognize alternating AT sequences should be studied.