Start/stop Codon-like Trinucleotides (CLTs) and Extended Clusters as New Language of DNA

Rosandić, Marija; Glunčić, Matko; Paar, Vladimir

Source / Izvornik: **Croatica Chemica Acta, 2011, 84, 334 - 341**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.5562/cca1948

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:097023

CROATICA CHEMICA ACTA
CCACAA, ISSN 0011-1643, e-ISSN 1334-417X
Croat. Chem. Acta **84** (2) (2011) 331–341.
CCA-3479

Preliminary Communication

Start/stop Codon-like Trinucleotides (CLTs) and Extended Clusters as New Language of DNA†

Marija Rosandić, Matko Glunčić, and Vladimir Paar*

Faculty of Science, University of Zagreb, Bijenička c. 32, HR-10000 Zagreb, Croatia

RECEIVED JULY 21, 2011; REVISED SEPTEMBER 15, 2011; ACCEPTED SEPTEMBER 16, 2011

Abstract. DNA nucleotide sequences carry genetic information of different kinds, not just coding instructions for protein synthesis. They can play a role, for example, in alternative conformations and gene regulators. The present paper introduces the extended start/stop codon-like trinucleotides (CLTs) for noncoding DNA sequences, based on trinucleotide cluster extension generated by specific single-nucleotide multiplications. Extended cluster analysis gives rise to rich information potential as a "new language" of DNA ("CLT-language"). The analysis of start/stop-CLTs extended clusters provides qualitative and quantitative differentiation and characterization of alpha satellites, as well as of other repetitive and non-repetitive noncoding sequences. As a measure of CLT extension of DNA sequences the extension factor *r* is introduced. Start/stop CLTs enable a distinction of three segments within alpha satellite, the first and the second as wrapping sequences and the third as a linker. Within a linker there are no start/stop CLTs. On the basis of start/stop-CLTs, it is hypothesized that these noncoding sequences may be involved in the networks of gene regulators. (doi: 10.5562/cca1948)

Keywords: alpha satellites, start/stop codons, nucleosome, cluster extension, gene regulators

INTRODUCTION

The role of noncoding sequences in expressing information has received much attention, in particular as gene regulators and carriers of several messages simultaneously.[1–8] On the other hand, much attention was given to alpha satellites, as intriguing noncoding arrays.[9–14] Alpha satellites are approximately periodic tandems of ~171 bp monomers in centromeric/pericentromeric regions of human and other mammal chromosomes. They could be hierarchically organized into higher order periodicity patterns known as higher-order repeat (HOR) alpha satellites. The HOR copies diverge from each other by less than 5 %, while alpha satellite copies within any HOR copy diverge from each other by ~20–35 % and divergence between alpha satellites outside of HORs is 20–40 %.[13] Unequal crossing over, restricted to tandem sequences, explains the generation and local homogenization of HOR units and accounts for large size variation among HORs on homologous chromosomes. HORs are particularly interesting since they, due to more recent evolution and by the process of unequal crossing over, enabled a rapid evolutionary progress. The available sequencing of centromeric region is still incomplete and HORs of human chromosoms are not completely sequenced.

METHODS

Identification and analysis of repeats and HORs and their consensus sequences is performed using the GRM (Global Repeat Map) algorithm[31–33] applied to the Build 37.2 and Build 2.2 assemblies of human and chimpanzee genomic sequences, respectively. GRM gives a direct mapping of symbolic DNA sequences into frequency domain providing a global map of repeats. An extended codon-like trinucleotide (CLT) analysis of genomic sequences is performed using the ECLT algorithm developed in this work. Full Methods and GRM and ECLT algorithms are freely available at request.

RESULTS AND DISCUSSION

In this paper a new approach of codon-like trinucleotides (CLTs) extensions, in particular for noncoding analogs of start/stop codons, is introduced to analyze the structure of noncoding sequences. This unusual viewpoint of terminators in noncoding sequences, in terms of CLTs, is resulting from detailed analysis of dinucleotides and trinucleotides in alpha satellites, presented in this paper. As a case study, this new method is applied to the following DNA sequences: to human and chim-

† This article belongs to the Special Issue *Chemistry of Living Systems* devoted to the intersection of chemistry with life.

* Author to whom correspondence should be addressed. (E-mail: paar@hazu.hr)

panzee alpha satellite consensus HORs, to monomeric alpha satellite regions and to no-repeat DNA regions (Table 1 and Supplementary table S1). In these studies the consensus sequences are determined by GRM analysis[25–33] of the recent Build 37.2 assembly (Figure 1).

First the dinucleotide frequencies in consensus alpha satellites are analyzed (Supplementary table S2). Previously, the pattern of dinucleotides and trinucleotides in noncoding sequences was extensively studied.[15–24] Two classes of alpha satellites are shown here: the poly-T class (with dominant TT dinucleotides), and the poly-A class (with dominant AA). It is seen here that all consensus HOR alpha satellites are of poly-T class, while the monomeric alpha satellites are of poly-A or poly-T class. Dinucleotide frequencies in human alpha satellite consensus HORs are shown in Figure 2. Earlier investigations have found a low frequency of TA dinucleotides in large segments of noncoding DNA. Here it is found that the TA-low pattern is most pronounced in HOR alpha satellites (Figure 2), which is not expected in light of A,T-rich character of alpha satellites. On the other hand, in no-repeat noncoding sequences the TA-low effect is much less pronounced.

Analysis of frequencies of trinucleotides in alpha satellite HORs is performed for consensus HORs in human chromosomes 1, 4, 5, 7, 8, 9, 10, 11, 17, 19, X, and Y (Table 1 and Supplementary tables S1 and S3). As an illustration, ten most frequent trinucleotides in 16mer alpha satellite HORs in human chromosome 7 are shown in Table 2, in comparison to previously computed ten most frequent trinucleotides in the genome of *C. elegans*[34] as comparison between evolutionary distant genomes. Out of ten most frequent trinucleotides in 16mer HOR in human chromosome 7, six are present in the genome of *C. elegans* too. The remainnig four trinucleotides in these sequences differ.

Different human and chimpanzee sequences from Table 1 are characterized by domination of TTT (poly-T class) or of AAA (poly-A class).

Inspired by much interest in gene regulators,[4–7] here the trinucleotides corresponding to the start (ATG) and stop (TGA, TAG, TAA) codons (CLTs) are analyzed in noncoding sequences. In each alpha satellite the two to four start- and ten to fourteen stop-CLTs are found (Figure 3 and Supplementary figure 1). The largest frequency is associated with the stop-TGA CLT (seven to ten). The main result of this study is the discovery that the CLTs are extended to cluster organization. This is illustrated in Table 3a for human consensus HORs.

In all alpha satellite monomers, the start-ATG CLT appears at two positions, separating monomers into two nearly equal parts. The first start-ATG CLT has always one A nucleotide overlap with the following stop-TGA CLT, thus forming the segment TGATG (Figure 3). It could be related to fusion of the last nucleotide of one alpha satellite monomer with the starting nucleotide of the next. The second start-ATG CLT is extended and overlapped to form ATTTGGA**,** representing a fusion of ATTTG and TTTGGA originating from extended start-ATG and stop-TGA CLTs. The extended stop-TAA CLT appears in all HOR alpha satellites only once, at the same position, as extended TAAAAA. The only stop-CLT which appears at three positions in alpha satellites in non-extended form is TAG.

It is found here that the stop-TGA CLT extension in sex chromosomes, human X and Y and chimpanzee Y, is twice smaller than in most of somatic chromosomes. Similarly, the investigated somatic chromosome 7 has by a factor of two reduced extension with respect to most of other human somatic chromosomes. Additionally, in alpha satellite HOR from human Y chromosome and in non-HOR poly-T alpha satellites in pericentromeric building blocks from chromosome 7 there appears the third start-ATG CLT between the two previous stop-CLTs (Figure 3). This contributes to a difference between human and chimpanzee Y chromosomes (Table 1).

It should be noted that small and seemingly insignificant differences in a nonlinear system of genes and regulators, as for example one additional start-ATG CLT in human Y chromosome with respect to chimpanzee Y chromosome, could produce significant functional differences.

Among the other CLTs, the largest frequencies and extensions are associated with CTT and TCC, but still by a factor of two smaller than the dominant extension of stop-TGA CLT. Extended cluster analysis shows a high degree of specificity for each alpha satellite monomer. Such a rich potential of specification enabled by rather high divergence between alpha satellite monomers of ~20−40 %, generated by the mechanism of extended CLTs, reveals a deeper structural organization.

All analyzed sequences have similar summary percentages of nucleotides forming start/stop-CLTs and percentages of extended start/stop-CLTs, embedded into ~60 bp C-free subsequence of alpha satellites (Table 1). This reveals a dominant influence of A,T-rich pattern. The TA dinucleotide, due to small frequency, has no significant impact on the results. Accordingly, the A,T-rich and TA-poor patterns do not show any differentiation between sequences, including both HOR and monomeric alpha satellites.

What significantly distinguishes these alpha satellite sequences from non-alpha satellite HORs and no-repeat sequences are the specificity and the level of extension of start/stop-CLTs, with dominant contribution from stop-TGA CLT (Table 1 and 3a). The largest extension of stop-TGA CLT in alpha satellites within

Table 1. Extended and non-extended stop-TGA and summary start/stop-CLTs in selected repeat and no-repeat sequences in human and chimpanzee chromosomes, expressed in percentages rounded off at the closest integer. Chromosomes without specification of species are human, while chimpanzee chromosomes are denoted by the notation chimp. The corresponding results for start-ATG, stop-TAA, and stop-TAG CLTs are given in Table S1

Sequence				Start/Stop CLTs		TGA				
Chr.	Structure	P.R.U.[a]	Poly-A/ Poly-T[b]	nt all E & NE[c]	all E[d]	nt E & NE[e]	nt E[f]	nt NE[g]	r[h]	r-class[i]
1	11mer HOR	alpha[j]	T	41	66	23	19	4	5	II
4	13mer HOR	alpha	T	37	71	22	20	1	17	I
5	13mer HOR	alpha	T	42	66	24	22	2	12	I
7	16mer HOR	alpha	T	43	69	23	22	2	14	I
8	11mer HOR	alpha	T	41	63	24	22	2	11	I
9	7mer HOR	alpha	T	40	62	22	21	2	14	I
10	18mer HOR	alpha	T	37	71	15	13	1	11	I
11	12mer HOR	alpha	T	40	68	21	18	3	7	II
17	14mer HOR	alpha	T	37	61	21	18	3	6	II
19	13mer HOR	alpha	T	41	65	24	22	2	13	I
19	17mer HOR	alpha	T	43	68	23	22	2	13	I
X	12mer HOR	alpha	T	41	66	23	20	3	7	II
Y	45mer HOR	alpha	T	44	71	25	22	3	8	II
chimp Y	30mer HOR	alpha	T	46	69	26	23	3	8	II
chimp 7	mon.[k]	alpha	T	48	68	25	21	4	5	II
7	mon.	alpha	T	40	71	21	20	1	17	I
7	mon.	alpha	T	47	70	24	23	2	15	I
7	mon.	alpha	A	44	65	20	16	4	5	II
7	mon.	alpha	A	44	65	20	16	4	5	II
1	mon.	alpha	A	51	62	23	19	4	5	II
11	mon.	alpha	T	46	69	24	20	4	5	II
Y	3mer HOR	1.6 kb	A	36	59	12	8	4	2	III
Y	5mer HOR	2.4 kb	T	48	63	12	10	2	5	II
Y	3mer HOR	0.55 kb	T	52	64	13	11	2	7	II
5	no-repeat[l]	-	T	45	64	11	9	3	3	III
21	no-repeat	-	A	47	63	12	9	3	3	III
7	no-repeat	-	T	44	63	12	9	3	3	III
1	random[m]	-	T	41	60	10	7	3	2	III
7	random	-	T	44	58	10	7	3	3	III

[a] Primary repeat unit.
[b] Type of sequence: A (Poly-A type), T (Poly-T type).
[c] Percentage of nucleotides in all extended and non-extended start/stop-CLTs in genomic sequence.
[d] Percentage of extended start/stop-CLTs with respect to all extended and non-extended start/stop-CLTs in genomic sequence.
[e] Percentage of nucleotides in extended and non-extended stop-TGA CLTs in genomic sequence.
[f] Percentage of nucleotides in extended stop-TGA CLTs in genomic sequence.
[g] Percentage of nucleotides in non-extended stop-TGA CLTs in genomic sequence.
[h] Quotient of nt E and nt NE.
[i] Classification of genomic sequence according to the r-value into three classes (see the text).
[j] Alpha satellite monomer.
[k] Monomeric alpha satellites.
[l] Illustrative 21 kb genomic sequences without repeats from human chromosomes 5 (contig NW_003315920.1, 23100–44100 bp), chromosome 7 (contig NW_003315922.1, 1–21000 bp), and chromosome 21 (contig NT_113952.1, 89250–110250 bp).
[m] Illustrative random sequences with nucleotide abundances the same as in alpha satellite HORs in chromosomes 1 and 7.

Table 2. Ten most frequent trinucleotides in consensus 16mer HOR in human chromosome 7 and comparison to the computation[34] of ten most frequent nucleotides in the genome of *C. elegans*

Human HOR in chr. 7		*C. elegans* genome[34]	
Trinucleotide	Frequency[(a)]	Trinucleotide	Frequency[(a)]
TTT	100	AAA	100
AAA	74	TTT	100
GAA	71	ATT	60
CTT	69	AAT	60
TTG	61	GAA	45
TCT	58	TTC	45
TTC	54	CAA	40
TGA	50	TTG	40
AGA	48	TCA	36
AAC	46	TGA	36

[(a)] Expressed relatively to the most frequent trinucleotide (normalized to 100).

HORs and in monomeric alpha satellites is mostly due to selective successive multiplications of T nucleotides (Table 3b), and less due to the multiplications of A nucleotides, resulting in poly-T or less frequently in poly-A alpha satellites. In HORs with poly-T consensus there are also some rare cases of individual poly-A alpha satellites.

The extensions of start-ATG and stop-TAA CLTs are significantly smaller than extensions of stop-TGA CLT, and the extension of stop-TAG CLT is absent or very small in all monomeric and HOR alpha satellites (Supplementary table S1). The stop-TAA CLT has a significant extension, but is of very low frequency. On the other hand, the extension of stop-TAG CLT is sizably larger in some non-alpha satellite HORs, which have primary repeat units of 1.6, 2.4, and 0.55 kb[32,33] sizably larger than alpha satellites (~0.171 kb).

As a measure of the CLT-extension of genomic sequences the corresponding extension factor r is introduced here (for definition of the factor r see caption to Table 1). Accordingly, the r-classification of genomic sequences is defined as follows:

class I for $r > 8$,
class II for $4 \le r \le 8$,
class III for $r < 4$.

Examples of these classes are (Table 1):
TGA class I: alpha satellite HOR in human chromosome 7,
TGA class II: alpha satellite HOR in human chromosome X,
TGA class III: no-repeat segment in human chromosome 21.

From these analyses the basic role of stop-TGA CLT is seen: the poly-T stretch preceding G nucleotide and the poly-A stretch following it (poly-T−G−poly-A) exhaust about 30 % of all T and A nucleotides in alpha satellites. The TA is bound to stop-TAA and stop-TAG (non-extended and extended) CLTs having small frequency and extension with respect to stop-TGA CLT. Only one TA is present in the 110 bp segments of alpha satellites outside of the start/stop-CLTs and their extensions. This results in TA-low pattern of alpha satellites (Figure 2).

In alpha satellite HORs from human chromosomes 1, 11 and 17 the extension of start/stop- CLTs is by a factor of two smaller than in other somatic human chromosomes, *i.e.*, similar as in human and chimpanzee sex chromosomes (Table 1). It could be hypothesized that in these segments of chromosomes the evolutionary changes were slower than in other somatic human chromosomes.

Chromosome 10 has a smaller number of stop-TGA CLTs, but their extension is similar as in other human somatic chromosomes.

Different multiplication patterns of individual A, T, G nucleotides in extension of start/stop-CLTs in alpha satellites might influence expression of codons in coding sequences. For example, stop-TGA CLT, which is most strongly extended within the alpha satellites, takes the form TTTGGAAA. We hypothesize that in this way it might exert the stop influence on some segments of the genome, possibly on codons underlying the extended stop-TGA CLT (Table 3c).

It is interesting to compare the concept of codon-like trinucleotides extension, introduced here, to the concept of Shannon N-gram extension used for reconstruction of the most likely sequence pattern by fusing overlapping triplets that was used for no-repeat sequences.[34,35] It was shown that Shannon's fusing can generate motifs like TTTTCGAAAA and TTTTTGAAAA[34,35] with a periodic structure on both sides of central motif, representing a nucleosome positioning patterns. Although a string similar to central motif starting with

Table 3. Illustrations of CLT extensions in alpha satellites. For description see the text.

a	TGA → TTGAA → TTTTGA → TTTTGAAA → TGGA → TTTGGA → TTTGA → TGGAAAA
b	T → TT → TTT → TTTT …
c	TTTGGAAA → TTT, TTG, TGG, GGA, GAA, AAA

chromosome 1 / 11mer α satellite HOR

```
m01 169 CTCAGAACTTCTTGGTGATGTTGGCATTCAACTGACAGAGTTGAACCTTCCCTTGTGAGTTCAGGTTGAAACGCTCTTTTCGTAGTATCTGCAAGTGGAGGTTTGGAACGCTTTGAGGCCTACGGTAGTAAAGGAAACAGCTTCATGTAAAAACTGGACAGAAGCATTC
m02 167 TCAGAAAATTCTTTGCGATGATTGAGTTTAACTCACAGAGCTGAGCATATCTTTTGATGGCGCATTTTCAAAACACACCTTTTGTAGAATATGCAAGTGGATTTTGGGACTTCTCTGAGAATTTCGTTGGAAACGGGATAAACCTCACATAACTGAAGAGGAACATT
m03 170 TCAGAACCTTCTTCGTGATGCTTGCATTCAACTCACAGTGTTGAACCTTTCTCTGACAGTTCAGGTTTGAAACACTCCTTCTGCAGAATCTGCAAGTGGAGATTTGGACCTCTTTGAGGCCTATCGTAGTAAAGGAAAGAACTTCATCTAAAAACAAGACAGAAGCATTC
m04 168 CCAGGAAACGCTTTGTGACGATTGAGTTCAACTCACAGAGCTGAACATTCCTTTGGATGGAGCAGTTTCCAAACACACTTTGTGTAGAATCTGCAAGTGGAGATTTGGACCGCTCTGAGGATTTCGTTGGATACGGGAGAAAAGTCACCTACGTAAACAGAAGCATTC
m05 171 TCAGAAGCTGCTTTGTGATGATTGCATTCAAGTCACCGAGTTGAACATTCCCTTTGATGGAGCCGTTTGGAAACACACTTTTGGTAGAATCTGAAAGGGGAGATTTGGACCGCTTTGAGGCCTATGGCAGTAGAGGATATAACTGCACATAAAAACGAGACAGGAGCATTC
m06 171 TCAGGAACTTCCTTGTGATGTTTGCATTCAACTCACAGAGTTGTACTTTCCTTTTGATAGAGCAGCTTTGAAACCCCCTCTTTCTAGCATCTGCAAGGGGACATTTGGAGGGCTTCGAGGCCTGGGGTGGAAAAGGAAATATCTTCTCATCAAAGCTACATGGAAGCATTC
m07 171 TCAGAAACTTCTTGGTGATGATTGCATTCAACTCACGGAGCTGAGGATTCCTTTTGATGCAGCAGTTTGGAACACTTCTTGTTGGAAATCTGCAAGCGGATATGTGGACCTCTTTGAACATTTCGATGGAAAAGGGATAATCTTCCCGTAAAAGCTAAACGGAAGCATGC
m08 171 TCAGAAACTCCTTTGTGAGGCTTGTGTTCAACTCCCAGAGTATAACATTGCTTTTCATAGAGCAGTTTTGAAACATTCTTTTCGTAGAGCCTCCAAGTGGACATTTGGAGCGCTTTCAGGCCTGTGGTGGAAAAGGAAATATCTTCACATAAAAACTAGAGAGAAGCATTG
m09 171 TCAGAAACTTCTCTGTGATGATTGCATGCAACTCACAGAGTTGAACATTCCTTTGATAGAGCAGTTTTGAAACTCTCTTTTGCTAGCATCTGCAAATGGATAGGTGGAACTCTGTGAAGACTTCTTTGGAAACGGGAATATCCTCACGTAAAAAGTAAACAGAAGCATTC
m10 170 GCAGAAACATCCTTCTGACGTTGGCATTCAAAGTCCAGAGTTCCTTTGGTAGTTCACGTTTGAAACACTCTTTTTGGAGGACCTGCAAGTGGATATTTGGAGCACTTTGTGGCCTTCGTTCGAAACGGCTATATCTTCACATAAAATCTAGACGAAGCCTTC
m11 167 TCAGAAAATACTTTGGGATGATTGAGTTCAACTCACAGAGCTGAACATTCCTTTGGGTGGAGCAGTTTTGAAACACACTTTTTGTAGACTCTGCAGGTGGATATTTGGACCTCTCTGAGGATTTCGTTGGAAACGGGATAACGTCACCTAACTAAACAGAAGCTTTC
```

chromosome 4 / 13mer α satellite HOR

```
m01 171 TCAGAAACTGCTTTGGGATGTGTGCATTCAACTCACAGAGTTGAACGCTTTCTTTGATACAGCAGTTTAGAAGCATTCTTTTTGTAGAATCTGCAAGTGTTCATTTGGAGCGCTTTGAAGCCTACGGTGGAAAAGGAAATATCTTCACATAAAAGCTAGACAGAAGTATTC
m02 172 TCAGAAACTGCTTTGTGATGTGTCCATTGAACTCACAGAGTCCAACCTTTCTTTTGATAGAGCAGATTTGAGGCATTCTCTGTGCATAATCTGCAAGTGTTCCTTTGGAGCGCTTTGAAGGCTTATGGTGGAAAGCAAAATATCTTCACATAAAAACTAAACAGAAGCATTC
m03 171 TCAGAAACTGCTCTGTGTTGTGTTCATTGAAGTCAAAGGGTTGAACTCCGTCTTTGATAGAGCAGTTTTGAAACATTCTTTTTGTAGAATCTGCAAGTGTTCATTTGGAGCGCTTTGAAGCCTAGGGTGCAAAAGGAAATATCTTCACAGAAAAACCGGACAGAAGCATTC
m04 171 TCAGACAGTTCCTTGTGATGTGCGCATTGAACTCAAAGAGTTGAACCTTCCTTTTGAGAAAGCAGTTTTGAAACAGTGTTTTCTTACGATCTACAAGTGGATATTTGGAGCGATTTGAGGCCTATGGTGGAAAAGGAAATATCTTCGGATAAAAACTACACAGAAGCATTC
m05 171 TCAGAAACTGCTTTGGGATGTGTGCATTCAACTCACAGAGGTGAACGCTTTCTTTGATAGAGCAGTTGTGAAACATTCTTGTTGCAGTATCTGCAAGTGTTCTTTTGGAGCGCTTCGAAACCTGCGGTGGAAATGGTAATATCGTCACATAAAAACTAGACGGAAGTTTGC
m06 171 TCAGAAACTTTTTTGTGATGTCTGCATTCAACTCACAGAGTTGAACCTTCCTTTGGGAGAGCAGCTCTGAAACAGTGTTTTTGCAGTATCACCCAAGTGGATATTTGGAGCGATATGAGGCCTATGGTGGAAAAGGAGATACCTTCACATGGAAGCTACACAGAAGCATTC
m07 169 TCAGAAACTTTTTTGGGATGTGTGGATTCAACTCACAGAGTTCACCCTTCCTTTTGATAGAGCAGTCTTGAAATACTGCTTTTGTAGAATCTGCTTGAGGATATTTGGAGCTCTTGAGGAATTCGTTGTAAACGGGGTATCTGCACATACAAACTGGACGGAAGCATTC
m08 170 TCAGAGACTTTTTTGTGATGTGTGAACTCAACTCACAGAGTTCAACCTTTCTTTTGATAGAGCAGTTTTGCAGCACTGCTTTTGTAGAATCTGCCTGAGGACATTTGGAGCTCTTTGAGGAATTCGTTGTAAACGGGATGTCCGCACATACAAACTAGACACAACTATTC
m09 170 TCTCAGAAAATTTTGTGTTGTGTGCATTCAAGTCACAGAGTTGAACGTCTTTTTGGATAGATCAGTTTTGAAACCATCCTTTTCTAGAATCAGCTTGAGGATATTTGTAGGTCTTTGAGGAATTCGTTGTAAACGGGATATCTGCACATACAAGGTAGACAGAAACATGG
m10 168 CCAGAAACTTTTTCGTGATGTGTGGATTCAACTCACAGAATTCAACCTTCCTTTTGATAGAGCATTCCTGAAACCCTGCTTGGGTAGAATCTGCTTGAGGATATTTGGAGCTCTTGGAGGAATTCGTGGCAAACGAGATATCTCCACATACAATCTAGACAGAAGCAT
m11 171 TCAGAAACTTTTTTGGGATGTGTGGATTCAACTCACACAGTTCAACCTTCCTTTTGATAGAGCAGTCTTGAAATACTGCTTTTGAAGAATCTGCTTCAGGATATTTGGAGCTCTTTCAGGAATTCGTTGCCAACGGGATATCTGCACAAAATTGACAGAAGCATTC
m12 166 CTCAGAAACATTTAGGGATGTGTGAACTCAACTCACAGAGTTCAACTTTCCTTTTGATAGAGCTGTCTTGAAATCCTGCTTTTGCAGAATCTGCTTGAGGATATTTGGAGCTCTTTCAGGAATTCGTTGCCAACGGGATATCTGCACAAAATTGACAGAAGCATTC
m13 170 TCAGACAGTTCTTTGTGATATGTCCATTCAACTCACAGAGTTGAATCTTTCCTTTGATAGAGAAGATTGGAAACATTCTTTCTGCATACTCTGCAAGTGTTCATTTGGAGCGCTTAGAGGCTCATGGTGGAAAGTGAAATATCTTCACATATAAACTAGACAGAAGCATT
```

chromosome 5 / 13mer α satellite HOR

```
m01 171 TCAGCAACTTCTTTGTGACGTTTGCATTCATCTCACAGTGTTGAACATACCTTTCCATAGAGTAGTTTTGAGACACTATTTTTGTAGAATCTGCAAGTGGATATTTGGACTGCTTTGAGGCCTTCATCGGAGACGGGAATATCTTCACATAAACACTAGACAGAAGCATTC
m02 170 TCAGAAACTTTTTTGTGATATGTTGATTCATCTGACAGCGTTGAACCTTTCTTTTGATAGAGCAGTTTTGAAAAACTCTTTTTGTCGAATCTGCAAGTAGACATTTGGAGTGCTTTGAGGGCTGTGGTGCCAAAGGAAATGTCTTCCCATGGAAACTAGACTGAAGCATTC
m03 170 TCAGAAAGTTCTTTGTGATGTGTGAATTCAACTCACAGAGTTGAACCTTCCTTTAATAGAGCAGTTTTGAAACACTCTTCTTCTAGAATCTGCAAGTAGATATTTGGAGCGCTTTGAGGCCTTCGTTGGAAACCGGAATATCTTCACAGAAAAAGTAGATAGAGGCATTC
m04 171 TCAGAAACTACTTTGTGATGTTTGCATTCAACTCACAGAGTTGAACATACCTCTTCATAGAGCAGTTTTGAAAACCTCTTTTTGTAGAATCTGCAAGTGGATATTCTGGACCACTTTGAGGCCTTCATAGGAAACAGTAATACCTTCACATAAAAACTAGATAGAAGCATTG
m05 171 TCACAATCGTCTTTGTGATGTGGGCATTCAACTAACACAGTTGAACATTTCTTCTCACAGAGCAGTTTTGAAACACTCTTTTGCTAGAATCTGCCAGTGGATACTTGGAGCGCTTTGAGGGCTATTGTGCCAATGGAGATATCTTCCCCTAAAAACTAGACAGAAGCATTC
m06 171 TCAGAAACTTCTTTGTGATGTGTGCGTTGAACCCAGAGAGATGAACCTTTCCTTTGATAGAGCAGTTTTGAAACGTGTTTTTGTAAGGTCTGCAAGCGGATAATGGGCTTCGCTTTGTGTCCTTTGGTGGAAACGGGAATATCTTCTAATAAAAACTAGACAGAAATATTC
m07 169 TCGGAAACATTTATGTGAAGCGTGCCTTCAACTCACAGAGTTGAACCTTCCTTTTGATAGAACAGTTTTGAAACACTCTTTTGAACAATTGCAGGTGAATCTTTGGAGCGCTTTGAAGCCTTTGTTGGAAATGGGAATATCTTCACACACAAACTAGCCAGAAGCATTC
m08 171 ACAGAAAATGCTTTGTGATGTGTGCATTCAAATCACGGAGTTGAATCTTTCTTTTGTTAGAGCAGTTTTGAAACACTGTTTCTGTGGAATCTGCCAGCGGACACTTGGAGCGCTTTGAGGGCTATGGTGGAGAAGGAAACATCTTCCCATAAAAACTAGAAAGAAGCATTC
m09 171 TCAGAAACTTCTTCGTGATGTGTGCATTCTACTCCCAAATTTGAATCTTCCTTCTCATGAAGCAGTTTTGAAACACTCTATTTGTGCAATCTACAATTGGATAATTGGAACCCTTTGATGCCCATGGTAGAAAAGGAAATATCCTCATATAAAAACTAGACAGAAGGATTC
m10 171 TCAGAAACTTCTTTAGGATGTCTGCAGTAAACTCACAGAGTTGAACATACCTTTCCGTAGAGCAGTTTTGAAACACTCTGTTTGTGGGATCCGCAAGTGGATATTTGGACAGCTTTGAGATCTTTGCTGGAAATGGGAATATCTTCACATATAAACTAGACAGAAGCATTC
m11 166 TCAGAAACTGCTTTGTGATATGTGCATTCGACTCACCGAGTTGAAACTTTTTTTTGATAGAGCAGTTTTGAAACACTCTGTAGAATCTGAAAGTGGATATTTGGAGCTCTTTGAGGGCTATGGCGGAAAAGAAAATATATTCACATTAAACTAGACAGCAGCATTC
m12 171 TGAGAAACTTCTTTGTGATGTTTGCATTCAACTACCAGAGTTGAATCTTCCTTTTGATAGGGCAGTTTGGAAACACTCTTTTTGTAGAATCTGCATGTGGATATCTGGAGCGATTTGAGGCCTATGGTCAAAAAGGAAATATCTTCCTGGGAAAAATAGACGAAAGCATTC
m13 171 TCAGAAACTACTTTGTGATCTGTCCATTCAACTCACAGAGTTGAACCTTCCTTTTTATGGAGCAGTTTTGAAACACTGTTTTTGGAGAATCTGCAAGTGGATATTTGGAGCGCTTTGAGGCCTATGGTAGAAAAAGAAATATCTGCCTCTAAAAACTAGACAGAAGCATTC
```

chromosome 7 / 16mer α satellite HOR

```
m01 171 TCAGAAACGAATTTGTGTTGTGTGCATTCTACTCCCATAGTTGAAAATTTCTTTTGATAGAGCAGTCTGGAACCACTCTGTTTCTAAAATCTGCAAATGGACATTTGGAGCGCTTTGAAGGTTATGATGGAAAAGGGAATATCTTCGCATTAAAACTAGACAGAAGCATTC
m02 171 TCAGAAACTTCTTTGTGATGTGTGCATTCAACTCCCAGGTTGAACCTTTCTTTTGTTAGAGCAGTTTTGAAACACTCCTTTTGTAGAATCTGCAGGCGGATATTTAAGTACTCTTTGAAGCATTCTTTGGAAACGAGAATATCTTCACCTAAAACCTAGACAGAAGCATTC
m03 172 TCAGAAACATCTTTGTGATGTGTCCATTCATCTCACAGAGTTGATAGAACAGTTTTGATAGAGCAGTTTTGAAACACTCTTTTTAAAGAATCTGCCAGTTCATATGTGCAGTGCTTTGAGGCTTATGGTAGAAAAGGAAATATCTTCATATAAAAACTAGACAGAAGCATTC
m04 171 TCAGAAACGACTTTGTGATGTGTGCATTCTACACACAAAGTTGAAACTTTCTTTTGATAGAGCAGTTTTGAAACCGTCTTTCCGAAGAATCTTCAAGTGGGCATTTCGAGGGCTTTGAGGACCATTGCGGATAAGGAAATATCTTCCCATAAGAAGTAGACAGAAGTATAA
m05 171 TCAGAAACTTCATTTTGATGTGTACATTCAACTCACAAAGCAGACCCTTACTTTTGATAGAGAAGTTTTGAAACACTCTTTTTGTAGAATCTGCAATTGGATATTTGGAGCGCTTTCAGGCCTCTGGTAGAAAAGGAAATATCTTCACATAAAAACTAGACAGAAGCATTC
m06 171 TCAGAAACGACTTTGTGATGTGTGTATTCTACTCCCATAGTTGAACATTTCTTTTGATAGAGCCGCCTGGAAACAATCTTCTTGTAGAATCTGCAAGTGGACATTTGGAGCGTTTCGAAGGCTGTGGTTGAAAAGGTAATATCTTCACCTAAAAACTAAATGGAAGCATTG
m07 170 TCCGAAACTTTTTGTGATGTGTGCGTTCAACTCACAGAGCTGAACCTTCCTTTTCATAGACCAGTTTTGAATCACTCTTTTTGTAGAATCCGCATTTAGATATTTGGAGCGCTTTGAAGACTTCATTGGAATCGCGAATACCTTCACATAAAAACTAGACAGAACCATTC
m08 171 TCAGAAACTTCTTTGAGATGTGTGCATTCAACTCACAGAGCTGAACCTTTCTTTTGATAGTGCAGTTTTGAAACATTCTTTTTTAAAAAATCTGCAGTTGGACATTTGGAGCTCTTTTAGGCTATCGGTTGAAAAGGAAATATCTTCACATTAAAACAAGACAGAAGCATTC
m09 170 TCAGAAACTCCTTTATGATGTCTGCATTCAACTCACAGAGTTGAACCTTCCTTTTCATAGAGCAGTTTTGAAACACTCTTTTCTGTAGAATCTGGAGGCGGATATTATCTTCACCTAAAAATTAGACAGAAGCATTC
m10 171 TCAGAAACTTCTTTGTGATGTGTGCATTCAACTCACAGCGTTGAAACTTCCTTTTGCCAGAGCAGTTTTGAAACCCTCTTTTTGAAGAATCTGAAAGTGCATAATTGCAGCACTTTGAGGCTTAAGGTCGAAAAGGAAATATCTTCATATAAAAACTAGACAGAAGCATTC
m11 171 TCAGAAACTACTTTGTGATGTGTGCATTCTACTCACATAGTTGAAATTTCCTTCTGATACTGCAGTTTTGAAACAGTCTTTTTGAGGGATCTTCAAGTGGGCATTTTGAGGGCTTTGGGGACTATTGTGGATAAGGAAATATCTTCACATGAAAAGTAGACAGAAGTGTTC
m12 170 TCAGAAACTTCATTTTGATGGGTGCATTCCACTAACAAAGTACAACCTTACTTTTATAGAGCAGTTTTGAAACAGTCTTTTTGTAGACTCTGCAAGTGGATATTTGGAGCGCTTTGAAGCCTTCGTTGGAAACGGGAATATCTTCCCCTTGAAACTAGACAGAAGCATTC
m13 170 TCAGAAACTTCTTTGTGATGTGGGCATTGAACTCACGGAGCTGAACCTTCCTTTGGATTGAGCAGTTTTGAAAAACTCTTCCTTTATAATCTGCAGGTGGATATTTGGAGTGCTTTGAAGCCTTCTTTGGAAACGGGAGTATCGTCACATAAAAATAGACAGAAGTATTC
m14 171 CCAGAAACTTCTTTGTGATTTGTCATTCAACTCACAGAGTTGAAGCTTCTTTTTGATAGAGCAGTTTTGAAACACCCTTTTTGCACAATCTGCAGGAGGATATTTGGAGCTCTTTGAGTGCTACATTGGAAACGGGAATATCGTCACCTAAAAACTAGAAAGAAGCATTC
m15 172 TCTGAAACCACTTTGTGATGTGTGCATTCATCTCACAGAGTTGAACCTTCCTTTTGATAGAGCAGTTTTGAAACCCTTTTTTGTACAATCTGCAAGTGGATATTTGGAGCAAATTGAAGCCTTCTTTGGAAATGGGAATATCTTAAATCTAAAAAATTAGGCAGAAGCATTC
m16 171 TCAGAAACTACTTTGTGATGTGTGCATTCAACTCACAGAATTGAACCTTCCTTTTGATACAGCAGTTTTGAAACACTCTTTGTTTAGAATCTGCAAGTGGATATTTGGAGCACATTTATGCCTGTGGTAGAAAAGGAAATATCTTCACATAAAAACTAGACAGAAGCATTC
```

Figure 1. Consensus HORs of human chromosomes obtained by using GRM algorithm[25,26] for computational analysis of Build 37.2 genomic assembly. Constituent alpha satellite monomers are aligned within each consensus HOR. Such alignment enables an easy insight into start/stop CLTs in alpha satellites.

chromosome 8 / 11mer α satellite HOR

```
m01 171 TCAGAACCTTGATTGTGATGTGTGTTCTCCACTAACAGAGTTGAACCTTTCTTTTGACAGAACTGTTTTGAAACATTCTTTTTATAGAATCTGGAAGTGGATATTTGGAAAGCTTTGAGGATTTCGTTGGAAACGGGAATATCTTCAAATAAAATCTAGCCAGAAGCATTC
m02 171 TAAGAAACATCTTAGGGATGTTTACATTCAAGTCACAGAGTTGAACATTCCCTTTCACAGAGCAGGTTTGAAACAATCTTCTCGTACTATCTGGAAGTGGACATTTTGAGCTCCTTGGGGCCTATGCTGAAAAAGGAAATATCTTCCGACAAAAACTAGACAGAAGCATTC
m03 171 GCAGAATCACGTTTGTGATGTGTGCACTCAACTGTCAGAATTGAACCTTGGTTTGGACAGAGCACTTTTGAAACACTCTTTTTTGTAGAATCTGCAGGTGGATATTTGGCTAGCTTTGAGGATTTCGTTGGAAACGGTAATGTCTTCAAAGAAAATCTAGACAGAAACATTC
m04 171 TCAGAAACACCTTCGTGATGTTTGCAATCAAGTCACAGAGTTGAACCTTCCGTTTCATAGAGCAGGTTGGAAACACTCTTTTTGTAGTATCTGGAAGTGGACATTTGGAGCGCTTTCAGGCCTATGGTGAAAAAGGAAATATCTTCCCATAAAAACGACATAGAAGCTATC
m05 165 TCAGGAACTTGTTTATGATGCATCTAATCAACTAACAGTGTTGAACCTTTGTACTGACAGAGCAGTTTGAAACACTCTTTTTTTGGAATCTGCAAGTGGATATTTGGATCGCTTTGAGGATTTCGTTGGAAACGGGATGCAATATAAAACGTACACAGCAGCATA
m06 171 TCAGAAAATACTTTGCCATATTTCCATTCAAGTCACAGAGTGGAACATTCCCATTCATAGAGCAGGTTTGAAACACTCTTTTTGGAGTATCTGGAAGTGGACATTTGGAGCGCTTTCTGAACTATGGTGAAAAAGGAAATATCTTCCAATGAAAACAAGACAGAAGCATTC
m07 167 TGAGAAACTTATTTGTGATGTGTGTCCTCAACTAACGGACTTGAACCTTTCGTTTCATGCAGTACTTCTGGAACACTCTTTTTGAAGATTCTGCATGCGGATATTTGGATAGCTTTGAGGATTTCGTTGGAAACGGGCTTACATATAAAAATTAGACAGCAGCATTC
m08 171 TCAGAAACTTCTTTGTGGTGTCTGCATTCAAGTCACAGAATTGAACATCCCCTCACATAGAGCAGTTGTGCAGCACTCTATTTGTAGTATCTCGAAGTGGACATTTGGAGGGCTTTGTAGCCTATCTGGAAAAAGGAAATATCTTCCCATGAATGCGAGATAGAAGTAATC
m09 167 TCAGAAACATGTTTATGCTGTATCTACTCAACTAACTGTGCTGAACATTTCTATTGATAGAGCAGTTTTGAGACACTCTTCTTTTGGAATCTGCAAGTGGATATTTGGATAGATTTGAGGATTTCGTTGGAAACGGGATTATATATCAAAAGTAGACAGCAGCATTC
m10 171 TCAGAAACTTCTTTGTGATGTTTGCATCCAGCTCTCAGAGTTGAACATTCCCTTTCATAGAGTAGGTTTGAAACCCTCTTTTTATAGTGTCTGGAAGCGGGCATTTGGAGCGCTTTCAGGCCTATGCTGAAAAAGGAAATATCTACCTACAGAAACTAGACAGAAGCATTC
m11 171 TGAGAATCACGTTTGTGATGTGGGTACTCAACTAACAGTGTTGATCCATTCTTTTGATACAGCAGTTTTGAACCACACTTTTTGTAGAATCTGCAAGTGGATATTTGGATAGCTGTGAGGATTTCGTTGGAAACGGGAATGTCTTCATAGAAAATTTAGACAGAAGCATTC
```

chromosome 9 / 7mer α satellite HOR

```
ch9b m01 170 TCAGAAACTTCCTTGTGATGTGTGTACTCAAGTAACAGAGTTGAACCTTACTTTTGACAGAGCCGTTTTGAAACAGTCTTTTTGTAGAATCTGGAAGTAGATATTTGGATACCTTTGAGGATTTCTTTGGAAACGGGATATCTTCATATAAAATCTAGACAGAAGCATTC
ch9b m02 171 TCAGAAACTTCTCTGTGATGTTTGCATTCAACTCATAGAGTTGAACACTTCCCTTCATAGAGCAGGTTTGAAACACTCTTTTTGTAATATTTGGAAGTGGACATTTGCAGCGCTTTGAGGCCTATGTTGAAAAAGGAAATATCTTCTCCTAAAAACCAGACAGAAGCATTC
ch9b m03 171 TCAGAAACTTATTTGCGATGTGTGTTCTCAACTAACAGAGTTGAACCTTTGTTTTGATACAGCATTTTGGAAACACTCTTTTTGTAGAATCTGCAGGTGGATATTTGGATAGCTTTGAAGGTTTCGTTGGAAACGGGAATATCTTCATATAAAATCAAGACAGAAGCATTC
ch9b m04 171 TCAGGAACTTCTTTGTGATGTTTGCATTCAAGTCACAGAACTGAACATTCCCTTTCATAGAGCATGTTTGAAACACTCTTTCTGTAGTATCTGCAAGCGGACGTTTCAAGCGCTTTCAGGCCTATGGTGAGAAAGGAAATATCTTCAAATAAAAACTAGACAGAAGCATTC
ch9b m05 167 TCAGAAACTTGTTTGTGATGTGTGTATTCAACTAACAGAGATGAACCTTTCTTTTTACAGAGCAGTTTTGAAACACTCTTTTTGTGGAATCTGAAAGTGGATATTTGGATAGCTTTGAGGATTTCGTTGGAAACGGGATTACATATAAAATCTAGAGAGAAGCATTC
ch9b m06 171 TCAGAAACTGCTTTGTGATGTTTTCATTCAAGTCACAGAGTAGAATGTTCCCTTTTATAGAGCAGGTTTGAAACACTCTTTCTGCACTACCTGGAAGTGGACATTTGGAGCGCTTTGAGGCCTATGTTGAAAAAGGAAATATCTTCCCATAAAAACTAGACAGAAGCATTC
ch9b m07 171 TCAGAAACTTCTTTGTGCTGTATGTCCTCAATTAACAGAGTTGAACCTTTGTGTGGATACAGCATTTTGGAAACATTCCTTTAGTAGAATCTGCAAGTTGATATTTAGATAGCTAGGAAGATTTCCTTGGAAACGGGAATATCTTCATATAAAATCTAGACGGAAGCATTC
```

chromosome 10 / 18mer α satellite HOR

```
m01 168 TCAGAAATTACTTTGTGATGTGTGCGTGCAACTCACGGAGATTAACCTTTCTTTTCATAGAGCAGTTTGGAAAGACTCTGTCTGTAAGGTCTGCAAGTGGATATTTAGATTTCTGTGAGGCCTTCGTTGCAAACGGGATTTCTTCATATACTCACAGACAGAAGAATT
m02 171 TCAGTAACTACTTTGTGTTGTGTGTATTCAACTCACAGATTTGAACCTTTCTTTAGACAGAGCAGATTTGAAACGCTCTTCTCGTGGCTTTTGCATGTGGAGGTTTCAAACGATTTGAGGCCAATGGTAGAAAAGGAAATATCTTCGTATAAAAACTAGAGAGAATCATTC
m03 169 TCAGAAACTACTTTGTGATGTGTGCGTTCCACTCACAGAGTTTAACCTTTCTTTTAATTGAGCAGTTTGGAAACACTATTTTTGTAAAGTCTGCAAGTGGATATTTGGACTTCTTTGAGCCCTTCGTTGGAAACGGGATTTCTCCATATACTGCTAGACAGAAGCATTT
m04 171 TCAGTCACTTCTTTGTGTTGTGTGCATTCAACTCAGAGATTTGAACCTTCCTTTAGAGAGAGCACATTTGAAACACTCTTTCTGTGTAATTTGCTAGTGCAGATTTCAAGCTCTTCGAGGACAATGGTAGAAAAGGAAATATCTTCGTATGAAAACTAGACAAAATCATTC
m05 169 TGAGATACTACCTTGTGATGTGTGCGTTCAACTCACAGAGTTTAACCTTTCTTTTCATAGAGCAGTTTGGAAACACTCTATTTGTAAAGTCTGCAAGTGGATATTTGGACCTCTTTGAGGCCTTCGTTGGAAACGGGATTTCTTCCTATAATGCTAGACAGAAGTATTC
m06 170 TCAGTAACTTGTTTGTGTTGTGTGTATTCAACTAACAGAGTTGAACCTTCCTTTAGAAAGAGCAGTTTTCAAACACTCTGTTTGTGCAATTTCCAATGGAGATTTCTAGGGATTTGAGGCCAGTCTTAGAAATGGAAATATCTTTGTATAAAAACTAGACAGTGTCATTC
m07 169 TCAGAAACTACTTTGTGATGTGTGCGTTCCACTCACAGACTTTAACCTTTCTTTTAATTGAGCAGTTTGGAAACACTCTCTTTGTAAAGTCTGCAGTAGGATATTTGGACCTCTTTGAGGCCTTCGTTGGAAACGGGATTTCTTCATATAATGCTAGATAGAAGAATTC
m08 171 CCAGTCACTTCTTTGTGTTGTGTGCATTCAACTCAGAGATTTGAACCTTCCTTTAGAGAGAGCACATTTGAAACACTCTTTTTGTGTAATTTGCTAGTGCAGATTTCAAGCTCTTCGAGGACAATGGTAGGAAAGGAAATATCTTTGTATTAAAACTAGACAAAATCATTC
m09 169 GCAGAAACTAGTTTGTGATGTGTGCGTTCAACTCACAGAGTTTAACGTTTCTTTTCATAGAGCAGTTTGGAAACGCTGTCTTTGTAAAGTCTGCAAGTGGATATTAGGACCTCTTTGAGGCCTTCGTTGGAAACGGGATTTCCTCCTATAATGCTAGACAGAAGAATTC
m10 171 TCAGTTACTACTTTCTGTTGTGTGCATTCAACTCACAGAGTTGAACCTTCCTATATTCAGAGCAGTTTTGAAACACTCTTTTTGTGGAATTTGCAAGTGGAGATTTCAAGGGATTTGAGGCCAATCTTAGAAATGGAAATATCTTCGAATTAAAACTACACACGAATCATTC
m11 169 TCAGAATCTACTTTGTGATGTGTGCGTGCAACTCACGGAGATTAACCTTTCTTTTCATAGAGAAGTTTGGAAAGAGTCTGTCTGTAAGGTCTGCAAGTGGATATTTAGATTTCTGTGAGGCCTTCGTCGCAAACGGGATTTCTTCATATACTGCCCGACAGAAGAATTC
m12 171 TCAGTAACTACTTTGTGTTGTGTGTATTCAACTCACAGATTTGAACCTTTCTTTAGACAGAGCAGATTTGAAACGCTCTTTTCGTGGCTTTTGCAAGTAAAGATTTCAAGCGATTTGAGGCCAATGGTAGAAAAGGAAATAGCTTCGTATAAAAACTAGACAGAATCATTC
m13 169 TCAGAAACTGCTCTGTGATTTGTGCGTTCAACTCACAGATTTTAAACTTTCTTTTCATAGAGCAGTTTGGAAACACTCTTTTTGTAAAGTCTGCAAGCGGATATTTGGACCTCTTTCAGGCCTTCTTTGGAAAAGGGATTTCTCCATATACTGCTAGCCCGAAGAATTT
m14 171 TCAGTAACTTCTTTGGGTTGTGTGTATTCAATTCACAGAGTTGAACCTTTCTTTAGACCGAGCAGATTTGAAACTCTCCTTTCGTTGCTTTTGCAAGTGGAGATTTCAAGCGATTTGAGGCCAATTGTAGAAAAGGAAATATCTTCGTATAAAAACTAGACAGAACAATTC
m15 169 TCAGAAACTTCTTTGTGATGTGTGCGTTCAACTCACAGAGTTTAACCTTTCTTTTCATAGAGCAGGTTGGAAGCACTCTCTTTGTAAAGTCTGCAAGCAGATATTTGGACCTTTTTGAGGCCTTCGTTGGAAACGGGATTTCTTCATATACTGCTAGACCAAAGAATTC
m16 171 TCAGTAACTTCTTTGTGCTGTGTGTATTCAACTCACAGAGTTGAACTTTTCTTTAGACAGAGCAGATTTGATACTCTCTTTTCGTGGCTTTTGCCAGAGGAGATTTCAAGTCATTGGAGGCCAATGGTAGAAAAGAAAATATCTTCGTATAATAACGAAACAGAATCATTC
m17 169 GCAGAAACTAGTTTGTGATGTGTGCGTTCAACTCACAGAGTTTAACGTTTCTTTTCATAGAGCAGTTTGGAAACGCTCTCTTTGTAAAGTCTCCAAGTGGATATTTGGAGCTCTTTGAGCCCTTCGTTGGAAACGGGACTTCTTCATATAATGCTAGACAGAAGAATAC
m18 171 CTCAGTAACTATTTGTGCTGTGTGCATTCAACTCACGGAGTTCAACCTTCCTTTATTCAGAGCAGTTTTGAAACACTCTTTTTGTGGAATTTGCAAGTGGAGATTTCAAGGGATTTGAGGCCAATCTTAGAAATGGAAATATCTTCGAATTAAAACAACACAGAATCGTTC
```

chromosome 11 / 12mer α satellite HOR

```
m01 171 TCGCAAACTTCTTTGTGATGTTTGTGTTCGGGTCACACAGTTTAACCTCGCTTTTCACAGAGCGGTTTTGAGACACTCCTTTCGTAGAATCTGCAAGTGGACATGTGGAGCGCTTCCAGGCCTGTGGTGGAAAAGGAAACATCTTCACATAAGAACTAGAGAGAAGCATTG
m02 171 TCAGAAACTTCTGTGTGATGATTGCATTCAGGTCACAGAGTTGAACCCTCCATTTGATTGAGCAGTTTGGAAACTCTCTTTTTGTAGAATCTGTAAGAGGATATGCGGACTTCTTTGAAGATTTCTTTGGAAACGGGAATATCTTCACAGAAAAACTAAACTGAAGCATTC
m03 170 TCCGAAACTTCTTTGTGATGTTTGCATACAACTCACAGAGTTGAACCTTCCTTTCATAGTTCAGCTTTGAGACACTCTTTTGGTAGAATCTGCAGGTGGATATTTGGACCACTGTGAGGCCTTCGTTCGAAACGGGTACACCTTCACGTAAAAACTCAAGAGGAGCATTC
m04 168 ACAGAAAATTCTTTGCGATGATTGTATTTAACTGAGAGAGCTGAACATTCGTTTGATGGAGCAGTTTCCAAACACACTTTTGGTAGAATCTGCAAGTGTAAATTTGGACTTCTCTGAGGATTGCGTTGGAAACGGGATAAACTTCCCAGAAGTAATCGGAAGCATTC
m05 171 TCAGAAACCACTTTGTGATGATTGCATCGGACTCACAGAGTTGGACATTCCTATGGATAGAACAGTTTGTAAACACTCTTTTTGTAGAATCTGCAATTGGAGATTTGGACGGCTTTGAGGCCTACGGAAGTAAAGGAAATAACTTCACATAAAAACCAAACGGAAGCATTC
m06 171 TCAGAAACTACTTTGTGATGTTTGTGTTCAACTTCCGGAATTGAACTTTCCTCTGGAAAGAGCAGCTATGAAACGCTCTTTTTCTAGAATGTGCAAGTGGACATTTGGAGGGCTTTGAGGCCTGCGGTGGAAAGGGAAATATCTTCACATGAAAACTAGATAGAAGCATTC
m07 171 TCAGAATGTTTCCTGTGAGGACTGCATTCAACTCACAGAGTTGAACAATCCTGTTGACGGAGCAGTTTTGAAACTCCCTTTCTTTGGAATCTGCAAGTGGATATGTGGACCTCTTTGAAGATTTCGTTGGAAACGGGTTCATCTTCACATAAAAACTAAACAGAAGCATTC
m08 169 TCCGAAACTTCTTTGTGATGTTTGCATTCAACTCACAGGCTGAACCTTCCTTTCATAGTTCAGCTTTCAAACACTCTTTCTGTAGAATCTGCAAGTGGATATTTGCACCACTTTGTGGCCTTCCTTCGAAACGGGTATATCTTCACATCAAACCTAGACAGAAGCATTC
m09 168 TCAGAAAATTCTTTGTGACGATGGAGTTTAACTCAGAGAGCTGAACATTCGTTTTGATGGAGCAGTTTCCAAACACACTTTTGGTAGAATCTGCAAGTGTAAATTTGGACTTCTCTAAGGATTTCGTTGGAAAGGGGATAAACTTCCCAGAAGTAATCGGAAGCATTC
m10 175 TCAGAAACGACTCTGTGATGATAGCATTCAACTCACAGAGTTGGACATTCATTCCTTTTGAGAGAGCAGTTTGGAAACACTCTTTCTGTCGAATCTGCAAGTGGAGATTTGGACCGCTTTGAGGCCTATGGTAGTAAAGGGAAGAACTTCATATAAGAACTAGACAGTAGCACTC
m11 171 TCAGAAACTTCTTTGTGATGTTTGCATTCACCTCACAGAGTCGAACTTTCCCTCTGATACAGCACCTTTGAAACGCTCGTTTTCTAGAATCTGCAGGTGGACATTGGGAGGGCTTTGTGGACTGTGGTGGAAAAGGGAATATCTTCTCATAAAAACTACATAGAAGCACTC
m12 171 TCAGAAACGTCTTTGTGATGATTGCATTCAACTCACAGAGTTGAAGATTCCGTTTGAAACAGCAGTTTCGAAACACTCTTTCTGTGGGATCGGCCAGTGGATATTTGGACCTCTTCGAAGATTTCGTTGGAAATGGGATAAACTTCACATAAAAGCTAAACCGAAGCATTC
```

Figure 1. Continued.

chromosome 17 / 14mer α satellite HOR

```
m01 171 TCAGAAACTACGTTGTGATGATTGCATTCACCTCACGGAGTGGAGCATTCCTATTGACAGAGCAGTTTGGAAACACTCTTGTTGTAGAATCTGCTAGTGGAGATTTGGAGCGCTTTGAGGCCTATGGTAGTAAAGGGAAGAGCTTCACATAAAATCTAGACAGAAGCATTC
m02 167 TCAGAAAATACTTTGTGATGATTGAGTTTAACACACAGAGCTGAACATTCCTTTGGATGGAGAAGGTTTGAAACACACTTTCTGTAGAATCTGCGAGTGGATATTTGGACCTCTCTGAGGATTTCGTTGGAAACGGGATAACTGCACCTAACTAAACGGAAGCATTC
m03 170 TCACAAAATTCTTTGTGATGTTTGCATTCAAATCCCAGAGTTGAACCTTCCTTTGATAGTTCAGCTTTGAAACACCCTTTTTGTAGGATCTGCAAGTGGATATTTGGACCACTCTGTTGGCCTTCGTTCGAAACGGGTACATCTTCAAATAAAATCTAGACAGAAGCCTTC
m04 171 TCAGAAACTTCTCTGGGATGATTGCATTCAACTCAAAGCGTTGAACCCTCCTATGGATAGAGCAGTTTTGAATCTCTCTTTTTGTGGAATCTGCAAGTGGATATGTGGTCCTCTTTGAAGATGTCTTTGGAAACGGGAATATCTTCACATAAAAACTAAACAGAAGCATTC
m05 171 TCAGAAACTTCTCTGTGATGTTTGTGTTCAAATCACAGAGTTTCACGTTGCTTTTCATAGAGCAGATGAGAAACATGCTTTTCGTAGGGTCTGCAAGTGGACATTTGGAGAGATTTCAGGCCTGTGGTGGAAAACGAATTATCGTCACGTAAAAACTAGAGAGAAGCATTG
m06 171 TCAGAAACTTGTTTGTGATGACTGCATTCAACTCACAGAGTTGAAGGTTCCTTTTCAAACAGCAGTTTCCAAACACTCTTTCTGTGGCATCTGCAAGTGGATGTTTGGGCCTCTTTGAAGATTTCGTTGGAAACGGGATAATCTTCACAGAAAAGCTAAACAGAAGCATTC
m07 171 TCAGAAACTTCTTTGTGATGTTTGCTTTCAACTCACAGAGTTGAACTTTCCTTTTGAGAGAGAAGCTTTGAAACACTCTTTTTCTAGAATCTGCAAGTGGATATTTGGAGGGCTTTGAGGCCTGTGGTGGAAAAGGAATTATCTTCCCGTAAGAACTAGATAGATGCATTC
m08 171 TCAGAAACTACTTTGTGACGATTGCATTCAAGTCACAGAGGTGAACATTCCCTTTCACAGAGCACTTTGGAAACTCTCGTTGTGTAGAATCTGCAAGTGGAGATATGGACCGCTTTGAGGCCTATGGTAGTAAAGGAAACAGCTTCATATAAAAACTAGACAGCAGCATTC
m09 168 TCAGAAAACTCTTTGTGACGACTGAGTTTAACTCACAGGGCTGAACATTCCTTTGGATGGAGCAGTTTGGAAACACACTATCTGTAGGATCTGCAAGCGGATACTTGGGCCTCTCTGAGGATTTCGTTGGAAACGGGATAAACCGCACAGAACTAAACAGAAGCATTC
m10 170 TCAGAACCTTCTTCGTGACGTTTGCATTCAACCCACAGTGTTGAACCTTTCTTTGATAGTTCAGGTTTGAAACACTCTTTTTGTAGAAACTGCAAGTGGATAACTGCACTTCTTTGAGGCCTATCGTAGTAAAGGAAATAACTTCCTATAAAAACAAGACAGAAGCTTTC
m11 167 TCAGAAAATTCTCTGGGATGATTGAGTTGAACTCACAGAGCAGTACTTTCCTTGGGATGGAGTAGTTTCGAAACACACTTTCTGTAGAATCTGCAAGTGGATATTTGGACCTGTCTGAGGAATTCGTTGCAAACGGGATAATTTCAGCTAAGTAAACAGAAGCAGTC
m12 169 TCAGAATCTTCTTGTGATGTTTGCATTCAAATCCCAGAATTGAACCTTCCTTTGAAAGTTCAGGTTGGAAACACTCTTTTTGCAGGATCTACAAGTGGATATTCGGACCACTCTGTGGACTTCGTTCGAAACGGGTATATCTTCACATAACATCTAGACAGAAGCATTC
m13 171 TCAGAAACTTTTCTGTGATGACTGCATTCAACTCACAGAGTTGAACACTCCTTTTGAGAGCGCAGTTTTGAAACTCTCTTTCTCTGGAATCTGCAAGGGGACATGCAGACCTCTTTGAAGGTTTCGTTGGAAACGGAATCATCTTCACATAAAAATTACACAGAAGCATTC
m14 171 TCAGGAACTCCTTGGTGATGTTTGTATTCAACTTCCAGAGTTGAACTTTCCTTCGGAAAAGAGCAGCTATGAAACACTCTTTTTCTAGAATCTGCAAGTGGACATTGGGAGGGCTGTGAGGTTTGTGGTGGAAAAGGAAATATCTCCACGTAAATACTAGATAGAAGCCTTC
```

chromosome 19 / 13mer α satellite HOR

```
m01 171 TCAGAAACTTCTTTGTGATCTGTCCATTCAACTCACAGAGTTGAACCTTCCTTTTTATGGAGCCGTTTTGAAACACTGTTTTTGTAGAATCTGCAAGTGGATATTTGGAGCGCTTTGAGGCCTAAGGTAGAAAAAGAAATATCTGCATATAAAAACTAGACAGAAGCATTC
m02 171 TCAGCAACTTCTCTTTGTGACGTTTGCATTCATCTCACAGTGTTGAACATACCTTTTCATAGAGCAGTTTTGAAACACTATTTTTGTAGAATCTGCAATTGGATATTTGGACTGCGTTGAGGCCTTCACTGGAAACGGGAATATCTTCACATAAACACTAGACAGAAGCATTC
m03 170 TCAGAAACTTTTTTGTGATATGTAGATTCAACTCACAGCGTTGAACCTTTCTTTTGATAGAGCAGTTTTGAAAAACTCTTTTATCGAATCTGCAAGTAGACATTTGGAGTGCTTTGAGGGCTGTGGTCCAAAAGGAAATATCTTCACATAGAAACTAGACTGAAGCATTC
m04 170 TCAGAAACTTCTTTGTGATGTGTGAATTCAACTCACAGAGTTGAACCTTCCTTTAATAGAGCAGTTTTGAAACACTCTTTTTGTAGAATCTGCAAGTAGATATTTGGAGCGCTTTGAGGCCTTCGTTGGAAACCGGAATGTCTTCACATAAAAAGTAGATAGAGGCATTC
m05 171 TCAGAAACTACTTTGTGATGTTTGCATTCAACTCACAGAGTTGAACATACCTCTTCATAGAGCAGTTTTGAAAACCTCTTTTTGTAGAATCTGCAAGTGGATATTTGGACCACTTTGAGGCCTTCATAGGAAACAGTAATATCTTCACATAAAAACTAGATGGAAGCATTC
m06 171 TCAGAATCTTCTTTGTGATGTGGGCATTCAACTAACACAGTTGAACATTTCTTTTGACAGAGCAGTTTTGAAACACTCTTTTAGTAGAATCTGCCAGTGGATATTTGGAGCGCTTTGAGGGCTATTGTGCAAATGGAAATATCTTCACCTAAAAACTAGACAGAAGCATTC
m07 170 TCAGAAACTTCTTTGTGATGTGTGCATTGAACCCAGAGAGATGAACCTTTCCTTTGATAGAGCAGTTTTGAAACGTGTTTTTGTAAGATCTGCAAGTGGATATTTGGGGCGCTTTGAGTCCTTTGGTGGAAACGGGAATATCTTCTAATAAAAACTAGACAGAAATATTC
m08 169 TCAGAAACATCTATGTGAAGTGTGCATTCAACTCACAGAGTTGAACCTTCCTTTTGATAGAACAGTTTTGAAACACTCTTTTGTACAATTGCAGGTGAATATTTGGAGCGCCTTGAAGCCTTTGTTGGAAATGGGAATATCTTCACATACAAACTAGCCAGAAGCATTC
m09 171 ACAGAAACTGCTTTGTGATGTGTGCATTCAAATCACGGAGTTGAACCTTTCTTTTGTTAGAGCAGTTTTGAAACACTGTTTCTGTGGAATCTGCCAGTGGACATTTGGAGCGCATTGAGGGCTATGGTGGAGAAGGAAATATCTTCACATAAAAACTAGAAAGAAGCATTC
m10 171 TCAGAAACTTCTTCGTGATGTGTGCATTCTACTCCCAAAGTTGAACCTTCCTTTTCATAAAGCAGTTTTGAAACACTCCTTTTGTACAATCTACAATTGGATAATTGGAACGCTTTGATGCCCGTGGTAGAAAAGGAAATCTCCTCATATAAAAACTAGACAGAAGGATTC
m11 171 TCAGAAACTTCTTTATGATGTTTGCATTAAACTCACAGAGTTGAACATACCTTTCCATAGAGCAGTTTTGAAACACTCTTTTTGTGGAATCCGCAAGTGGATATTTGGACCGCTTTGAGACCTTCGCTGGAAATGGGAATATCTTCACATATAAACTAGACAGAAGCATTC
m12 166 TCAGAAACTACTTTGTGTTATGTGCATTCAACTCACAGAGTAGAACCTTTTTTTTGATAGAGCAGTTTTGAAACACTCTGTAGAATCTGAAAGTGGATATTTGGAGCTCTTTGAGGGCTATGGTGGAAAAGAAAATATATTCACATTAAACTAGACAGAAGCATTC
m13 173 TCAGAAACTTCTTTGTGATGTTTGCATTCAACTACCAGAGTTGAACCTTCCTTTTGATAGAGCAGTTTTGAAACACTCTTTTTGTAGAATCTGCATGTGGATATCTGGAGCGATTTTTGAGGCCTACGGTCAAAAAGGAAATATCTTCCTATGAAAAATAGACAAAAGCATTC
```

chromosome 19 / 17mer α satellite HOR

```
m01 171 TGTAAAACTCCTTTGTGATGTTTGCATTCAACTCAGAAAGTTGAACTTCTCTTTATATAGTCCAGTTTTCAAACACTATTTTTGTAGAATCTGCAAGTGGATACTGGGACTGCTTTGAGGCCTTCGTTGGAAACGGGATTATCTTCACATAAAAACTAGACTGAAGGATTC
m02 171 TTAGAAACTTCTTTGTGATGTGTGCATTCAACTCACCGAGTGGAACCTCACTTTTGATAGAGCAGTGTTGAAAGACACTTGTTGTAGAATCTGCAGGTGGATATTTGGAGTGCTTTGAAGCCTTCCTTGGAAACGGGAATATCTTCACATAAAAACTAGACATAAGCATTC
m03 171 TCAGAAACTCCTTTGTGATCTGTCCATTCAGCTCACAGAGTTGAACCTTCCTTTTGATAGAGCAGTTTTGAAACACTCTTTCTGTAGAGTCTGCAAGTGGATATCAGGAGCGCTTTGAGGCCTATGGTAGAAAAAGAAATATCTGCATATAAAAACTAGACAGAAGCATTC
m04 171 TGAGAAACTTCTTTGTGATGTTTGCATTCAACTACCAGAGTTGAACCTTCCTTTTGATAGAGCAGTTTTGAAACACTCTTTGTGTAGAATCTGCATGTGGATATCTGGAGCGATTTGAGGCCTATGGTCAAAAAGGAAATATCTTCCTATGAAAAACAGACAAAAGCATTC
m05 166 TCAGAAACTACTTAGAGATATGTGCATTCAACTCACAGAGTTGAAACTTTTTTTTGATAGAGCAGTTTTGAAACACTCTGTAGAATCTGAAAGTGGATATTTGGAGCTATTTGAGGGCTATGGTGGAAAAGAAAATATATTCCCATTAAACTAGACAGAAGCATCC
m06 171 TCAGAAACTTCTTTATGATGTTTGCATTAAACTCACAGAGTTGAACATACCTTTCCATAGAGCAGTTTTGAAACACTCTTTTTGGGGAATCCGCAAGTGGATATTTGGACCGCTTTGAGACCTTTGCTGGAAATGGGAATATCTTCACATATAAACTAGACAGAAGCAATC
m07 170 TCGGAAACTTCTTCGTGATGTGTGCATTCTGCTCCCAAAGTTGAACCTTCCTTTTCATAAAGCAGTTTTGAAACACTCTTTTGTACAATCTACCATTGGATATGTGGAAGGCTTTGATGCCCATGGTAGAAAAGGAAACATCCTCATATAAAATCTAGACAGAAGGATTC
m08 171 ACAGAAACTGCTGTGTGATGTGTGCATCCAAATCACGGAGTTGAACTTTTCTTTTGTTAGAGCAGTTTTGAAACCCTGTTTCCGTGGAATCTGCCAGTGGACATTTGGAGCGCATTGAGGGCTATGGTGGAGAAGGAAATATCTTCACATAAAAACTAGAAAGAAGCATTC
m09 169 TCAGAAACATCTATGTGAAGTGTGCATTCAACTCACAGAGTTGAACCTTCCTTTTGATAGAAGAGTTTTGAAACACTCTTTTGTACAATTGCAGGTGAATATTTGGAGCGCTTTGTGAAGCCTTTGTTGGAAATGGGAATATCCTCACATAAAAACTAGCCAGAAGCATTC
m10 170 TCAGAAACTTCTTTGTGATGTGTGCATTGAACCCAGAGAGATGAACCGTTCCTTTGAGAGAGCAGTTTTGAAACGTGTTTTTGTAAGATCTGCAAGTGGATATTTGGGGCGCTTTGAGTCCTTAGGTGGAAACGGGAATATCTTCGAATAAAAACTAGACAGAATTATTC
m11 171 TCAGAATCTTCTTTGTGATGTGGGCATTCAACTAACACAGTTGAACATTTCTTTTGACAGAGCAGTTCTGAAACACTCTTTTTGTAGAATCCGCCAGTGGATATTTGGAGCGCTTTGAGGGCTATTGTGCAAATGGAAATATCTTCACCTAAAAACTAGACCGAAGCAATC
m12 171 CCAGAAACTACTTTGTGATGTTTGCATTCAACTCACAGAGTTGAACCTACCTCTTCATAGAGCAGTTTGGAAAACCTCTTTTTGTAGAATCTGCAAGTGGATATTTGGACCACTTTGAGGCCTTCATAGGAAACAGTACTATCTTCACATAAAAACTAGGTAGAAGCATTC
m13 170 TCAGAAACTTCTTTGTGATGTGTGAATTCAACTCACAGAGTTGAACCTTCCTTTAATAGAGCAGTTTTGAAACACTCTTTTTGTAGAATCTGCCAGTAGATATTTGGAGCGCTTTGAGGCCTTCGTTGGAAACCGGAATATCTTCACATAAAAAGTAGATAGAGGCATTC
m14 170 TCAGAAACTTTTTTGTGATATGTAGATTCAACTCACAGCGCTGAACCTTTCTTTTGATAGAGCAGTTTTGAAAAACTCTTTTATCGAATCTGCAAGTAGACATTTGGAGTGCTTTGAGGGCTGTGGTCGAAAAGGAAATATCTTCACATAGAAACTAGACTGAAGCATTC
m15 171 TCAGCAACTTCTTTGTGACGTTTGCATTCATCTCACAGTGTTGAACATACCTTTTCATAGAGCAGTTTTGAAACACTATTTTTGTAGTATCTGCAAGTGGATATTTGGACTGCTTTGAGGCCTTCATTGGAAACGGGAATATCTTCACATAAACACTAGACAGAAGCATTC
m16 170 TCTGAAACTTCTTTGTGATGTGTGTATTCAACTCACAGAGTTGAACCATCTTTTTTATGGAGCGGTTTTGAAACAGTGTTTTTGTAGAATCAGCAAGTGGATATTGGGAGCGCTTTGAGGCCTCTGGTGGAAAGGGAATGTCTTCACATAAAAACTGGACAGAAGCATTC
m17 171 TCAGAAACATCTTTGTGATGTTTGCATTCAACTCACAGAGTTGATCCTTCCTTTTAATAGGGCAGTTTTGCAACACTCTTTTTGTAGAATGCACCAGTGGGCTTTTGGAGCACGTCAAGGGCTATGGTGAAAAAGGAAATATCTTCACATAAAAACTAGACAGAAGTATTC
```

Figure 1. Continued.

chromosome X / 12mer α satellite HOR

```
n01 170 TCAGAACATTCTTTGTGATGTTTGTATTCAACTCACAGAGTTGAACCTTCCTTTGATAGTTCAGGTTTGCAACACCCTTGTAGTAGAATCTGCAAGTGTATATTTTGACCACTTTGTAGCCTTCGTTTGAAACGTCTATATCTTCACATCAAACCTAGACAGAAGCATTC
n02 168 TCAGAAAATTCTTTGTGACGATGGAGTTTAACTCAGAGAGCTGAACATTCGTTATGATGGAGCAGTTTCCAAACACACGTTTTGTAGAATCTGCAAGGGGATATTTGGACCTCTCTGAGGATTTCGTTGGAAACGGGATCAACTTCCCATAACTGAACGGAAGCAAAC
n03 171 TCAGAAACTGCTCTGTGATGATTGCATTCAACTCCCAGAGTTGAACATTCCTTTTGATAGAGCAGTTTGCAAACACTCTTTTTGTAGAATCTGCAAGTGGAGATTTGGACCGCTTGTGGTAGTAAAGGAAAGAACTTCATATAAAAACTAGACGGTAGCACTC
n04 169 TCAGAAACTTCTTTGGGATGTTTGCATTCACCTCACAGAGTTGAACTTTCCCTTTGATAGCGCAGCTTCGACACACTTTTTCTACAATGTGCAAGTGGATATTTAGCGGGCTTGGAGGACTGTGTTGGAAAAGGAAATATCTTCTCCTAAAAACGACATAGAAGCATTC
n05 171 TCAGAAACTTCTTTGTGATGATTGCATTCAACTCACAGAGTTGAAGATTCCTTTTGAAACAGCAGTTTCGAAACACTCTTTCTGTGGGATCCGCAAGGGGATATTTGGACCTCTTTGAAGATTTCGTTGGAAACGGGATAATCTTCACCTAAAAGCTAAACGGAAGCATTC
n06 171 TCAGAAACTGCTTTGTGATGTTTGTGTTCGAGCCGCAGAGTTTAACATTGCTTTTCATAGAGCAGTTTTGAAATATTCTTTTGGCAGAATCTGCAAGTGGACATTTGGAGCGCTTTCAGGCCTGTGGTGGAAAAGGCCTGAAAGCCTTTTCCTTTACTTCACAGAAAGAC
n07 171 TCAGAAACTTCTGAGTGATGATTGCATTCAAGTCACACAGTTGAACCCTCCTTTTGATTGAGCAGTTTTGAAACTGTCTTTTTGTAGAATCTGTAAGTGGATACGTGGACCTCTTTGAAGATTTCTTTGGAAACGGGAATATTTCCACAGAAAAACTAAACTGAAGCATTC
n08 170 TGAGAAAGTTCTTCATGATGAATGCATTTAACTCGCAGAGATGAACCTGCCTTTGAGAGTTCAGGTTCGAAACACTCTTTCTGTAGAATCTGCAAGTGGATATTTGGACCACTGGGTGGCCTTCGTTCGAAACGGGTATATGTTCACGTAAAAACTAAAGAGAAGCATTC
n09 168 TCAGAATATTCTTTGTGATGATGGAGTTTCACTCACAGAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAAATACACTTTTGGTAGAATCTGCAGGTGGATATTTGGACCTCTCTGAGGATTTCGTTGGAAACGGGAATAATTTCCCATAACTAAACACAAACACTC
n10 171 TCAGAAACGACTTTGTGAGGATGGCATTCAACTCATGGAGTTGAACAATCCTATTGATAGAGCAGATTGGAATCACTCTTTTTGTAGAATCTGCAAATGGAGATTTGGACTGCTTTGAGGCCTACGGTAGTATAGGAAGGAACTTCATATAAAAGGCAAACGGAAGCATTC
n11 171 TCGGAAACTACTTTGTGATGTTTGTATTCAACTCCCAGAGTTGAACTTTCCTTTTGAAAGAGCAGCTATGAAACACTCTTTTTCGAGAATCTGCAAGTGGACGTTTGGAGGGCTTTGAGGCCTGTGGTGGAAAAGGAAATATCTTCACATAAAAACTAGATAGAAGCATTC
n12 171 TCAGAAAGTTTTCTGCGATGACTGCATTCAACTCACAGAGTTGAACAATCCTTTTGATGGAGCAGTTTTGAAACCCTCTTTCTTTGGAATCTGCAAGGGGATATGTGGACCTCTTTGAAGATTTCACTGGAAACGGGATCATCTTCACATAAGAACTAAACAGAAGCATTC
```

chromosome Y / 45mer α satellite HOR

```
n01 171 TTCTGAGAAACTTCTTTGTGATGTTCTTAACCATCTCACAGAGTTGAACCTTTCTTTTGATTCGGCAGTTTTGAAACACGGTTTTTGTAGAATCTTCAAGTGGATATTTGGAGCACTTTTCTGCCTATTGTGTAAAAGGAAATATCTTTACGTAAGAACTACACAGAAGCA
n02 171 TGAGAAACTCCTTTGTTGATGTGTCATTCACCTCACAGATTTGAATGTCTCTTTGATTGAGCAGTTTGGAAGCACTCTTTCGGTAGAATCTGCAAGTGGATATGGAGAGAGCTTTGAGGCCTGTTGTGGAAAACTAAATGTCTTCATATAAAAGCTACACAGAAGCATTC
n03 167 TGAGAAACTTCTTTACGATGTCTGCATTCAACTCACAGAATTGAACCTTTTGTTTGAGTAGTTTTGAAACTCTCTTTTTGTAGAATCTAGAAGTGGATATTTAGAACGCTTGGAGGCCTATGGTGCAAAAACGAATAACTTCACACAAAAAATACACAGAAGCATTC
n04 168 TGAGAAACTTCTTTGTGATGTTTGCTTTCACCTCCGGGAGTTCAACCTATCATTTGACAGAGCGGTTTTGAAACTCTTTTTGTAGAATCTCCAAGTGGATATTTGGAGCCCTTTGCATTCTACTGTGAAAAGGAAATATCTTCACATCAAAACTACACAGACGCATTC
n05 171 TGAGAAACCTCTTTGTGATGTGTGTATTCATGTCACGGTTTTGAACCTTCCATTTGATTGAGCAGTTTTGAAATACTCGTTTGGTAGAATGTACAAGTGAATATTTGGAGCACTTTGAGGCCTATGATAGAAACGGAAATATGTTTACATAAAAACTACACAGAAGCATGC
n06 172 TGAGAAACTACTTTGTGATGTGGGCATTCATCTCACAGAGTTGAAATCATTTTTTGATTTGAGTAGTTTGGAAACACTCTTTTTGTGGAATCTCTAAGGGCATATTTGAAGCGTTTTGCACGCTGTTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGCATTC
n07 170 TGAGAAACTAGTTTGTGATGTGTGCATTCAACTCACAGAGACGAACTGATCTTTTGATAGAGCAGTTTTGAAACTCACTTTTGTAGAATCTGCAGGTGGATATTTGGAGTACATTGCGGCCTATGGTGAAAAAGGAACTATCTTCGCATGAGAACCAGGCAGAAACATTC
n08 171 TGAGAAGCTTCTTTTTGATGTATGCATTCATCTCACGGTGTTGAAACTTTATTTTTATTGAGCAATTTTGAACATTCCTTTTTATAGAATCTACAAGTGGATATTTGGAGTGGTTTGAGACCTATGGTAGAAAAAGAACTATCTTCACCGAAAAACCACACAGAAGCATTT
n09 171 TGAGAAACTCCTTTGTGATTTGTCACTCAACCCACAAAGTTGAACATATTGTTTGATAGAGCAGTTGTGAAACTCTCTTTTTGTAGAATCTGCAAGTGGGTATTTGGAGCCCTTTGTGGCCCATGGTAGAAAAGGAACTATCTTCACAGAAAAACTACCCAGAAGCATTT
n10 170 GGGAGAAAATTCTTGTGATATTTGTGTTCATCTCATATATTTGAACCTTTCTTTTCATTGTGCAGTTTCCAAGCAATCTTTTTCTAGAATATGTAAGTGGATATTTGGAGCACTTTGTGGACTATGGAGGGAAAAGAAATGTCTTCACATAAAAACTACACAGAAGCATT
n11 170 TGAGAAACTTCTTTGTGATGTGTGCATTCATCAAACAGAATTGAACATTTCTTTTTTTGTGCAGTTTTGAAACAATCTTCTTGTAGTATCTGCAAGTGGATATTTGGAGCGTTTTAAGACCTAAGGTGGGAAAGGAAATATCTTCACATAAAAATTACACAGAGAGATTC
n12 171 TGAGAAACTTCTTTGTGATGTGTGCATTCATCTCACAGAGTTGAACATTTCCTATGATTGAGCAGTTTGGAAATATTCTTTTCATAGAATCTGGAAGTGGATATTTGGAGCCCTTTGAGGCCTATTGTGGAAAAGGAAATATCTTCACATAAAAACTACAGAGAAGCATTC
n13 170 CTGAGAAATTTCTTGTGATGTGTGCATTCATCTCACAGAGTTGAACCTTTCTTTTGATTGAGCAGTTTTGAAATACTCCTTTTGTAGAATCTGCAAGTGGATATTTTGAGTGCTTTGAGAACTATTGTGGAAAAGGAATTATCTTCTCATAAAACCTACACTGAAGGATT
n14 170 TGAGAAACTTCTTTGTGATGTGTCCATTCAACTCACATAGTTGAAACCATCTCTTGATCGAGTAGTTTTGAACCTCTCTTGTTGTAGAATCTGAAAGTGGATATTTGTGTCCCCTGGCGGTCTATGGTGGAAAAGAAATATCTTCACAAAAATACTACACAGAAGCATTC
n15 171 CGACATAGTTCTTTGTGATGTGTGCATTCATCTCACCTAGTTGAACCTTTCTTTTGGTTGTGCACTTTTGAAACACTCTTTTTGTGGAATCTGCAAGTGGATATCTGGATCACTTTGACGTCTATTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGAATTC
n16 171 TGAGAAACTTCCTTCTGATGTGTGCATTCATCTCACAGAGTTAAATCTTTCTTTTGATTGAGCAGTTTGCAAACACTCTTTTTGTGGTATCTCCAGGAGGATATTTGGAGTGCTTTGAGGCCTATGTTGGAAAAGGAAGTATCTTCCCTTAAAAGCTATGCAGAAGCATTC
n17 171 TGAGAAACTTCTTTGTCATGTGTGGATTCTTCTCACAGAGTTGAACTATCTTTTGATTGAGCAGTTTTGAAACACTGTTTTTTTTAGAATCTGCAAGTGAATATTTGGAGCCTTTTGGGTCTTATTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGCATTC
n18 172 TGAGAAACTTCTCAGTGATGTGAGCATTCTTCTCACAGAGTTGAACCTTTCTTTTGATTGAGCAGTTTTGAAACACTGTTTTTTTTAGAATCTGCAAGTGAATATTTGGAGCCTTTTGGGTCTTATTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGCATTC
n19 171 TGAGAAACTTCTCAGTGATGTGAGCATTCAAACCACAGAGTTGAACCTATCTTTTGAATGAGCAGTTTTGAAACTCTCTTTTCATAGTATCTGCAAGTGGATATTTGGAGCCTTTTGTGGCCTACGGTGGGAAAGGAAATATCTTCATATAAAAACTACACAGAAGCATTC
n20 168 TGTGAAACTTCTTTGTGCTGTGTGCATTCACCTCACAGAGTTCAACTTTATTTGATACAGCAGTTTTGAAACACTCTTCTTGTGGAATCTGCAAGTGGAAATTGGGAAATATTTAGGCATATGGTGGAAAAGGAAACATCCGCACATAAAAACTACACAGACACATTC
n21 171 TGAGAAACCATTTTGTGCTGTGTGCATTAATTTCCCAGAGTCGAACCTTTCTTTTGATTGAGCAGTTTTGAAACACTCTTTTTGTAGAATCTGCAAGTGGACATTTGAAGCACTTTGAGGCCTATTGTTGAAAAGGAAACATCTTCATATAAAAACAACAAGGAAGCATTC
n22 167 TGAGAAACTTCTTTGTGATGTGTGCATTCATCTCACAGAGTTCAACCTATCTTTCGTAGAGCAGTTTTGAAACTCTCTTTTCCTAGAATCTGTAAGTTGATATTTGGAGCCCTTTGTGCGGCCTATTGTGGAAAAGGAAATAACTTCACATGAAAACTACACAGAAGC
n23 171 TGAGAAACTTCTTTGTGATGTGTGCATTCAAACCACAGACTTGAACTGATCTTTTGATAGAGCAGTTTTTAAAGTGTCTTTCTGTAGAATCTGCAAGTGGTTACTTGGAGACCTTTGTGGAAGATGGTGGAAAAGGAAATATCTTCCCGTAAAAACTACACAGATGCATTC
n24 171 TGAGAAAATTCTTTGTGATGTGTGCATTCATCACAAAGAGTTGTACCTTTCTTTTGATTGAGCAGTTTTGAAACACTCTTTTTGTAGAATCTGCAAGTGGATATTTGGAGCACTTTGAGGCCTATGGTGGAAAAGAAAATATCTTCATATGAAAACTACACAGAAGCATTC
n25 171 TGAGAAACGTCTTTGTTATTAGTGAATTCATCTCACAGGGTTGAACGTTTCATTTGATTGAGCAGTTTTGAAACACTCTTTTTGTAGAATCTGCTATTGGATATTTGTAGTGCTTTGATGCCTATGGTAGAAAAGGAAAGATCTTCACATAAAAACTACACAGAGAAGCATTG
n26 169 TGAGAAACTATCTTTTAATGTCTGCATTAATCTCACGGAGTTCAACCTATCTTTTGATTGAGGAGTTTTGAATCTCTTTTTTGTAGAATCTGCAAGTGGATACTTGGAGCCCTTTCCGGCCTATTTTGGAAAAGGAAATAAGTTCACATAAAAACTACCCAGAACCATAA
n27 171 TGAGAAACTTCGTTGTGATGTATGCATTCAACTCATGCAGCTGAACCTCTTTTTTTGACAGAGCAGTTTTGAAACTCTCATTTTGTAGAATCTGCAAGTGTATATTTGGATATCTTTGAGACATATAGTGGAAAAGGAAATATCTTCCCATAAAACTACAATGAAGCCCTC
n28 171 TGAGAAACTTCTTTGTGATGTGTGCATTCATCTCACAGAGTTGAGCCTTTCTTTTGATTGAGCAGTTTTGAAAATCTCTTTTTGTAGAATCTGCAAGTGGATATTTGGAGCCCTTTGCAGCCTATGGTGGAAAAGAAAATATCTTAACATAAAAAATACACAGAAGCATTC
n29 171 TGCAGAAACTCCTTTGTTTTGTGTGCATTCATTTCTTAGAGTTGAAGCTTCCATTTGATTGAGCAGTTTTGAAACACTCTTTTTGTAGAATCCACATTTGGATATTTGGAGGGCTTCGAGACCTATTGTGGAAAAGGAAGTATCTTCACTTAAAAACTACACAGAAGCATT
n30 171 TGAGAAACTTCTTTGTCAGGTGTGTATTCAACTCGCAGAGCTGAACCTATCTTTTGACAGAGCAGTTTTGAAACTCTCCTTGTGTAGAATCTGCATGTGGTTATTTGGAGGCCTTTGTGACCGATGGTGGAAAGGAAAACGCCTTCCCCAAAAAACTACACAAATCATTC
n31 171 TGAGAAACTTGTCTGTGATGTGTGCATTCATCTCACAGAGTTGAACTTTCTTTTGCTTGAGCAGTTTTGAAAGACTCTTTTTGTAGTTTCTGCAAGTGGATATTTGGAGTGCTTTGTGGCCTAAAGTGGAAAAGGAAATATCTTCACATGAAAACTACACAGAAGCATTC
n32 170 TCAGAAACTCCTTTGTGATGTGTGCATTCATCTCACAGTGTTGAAACATTCTTTTAACTGAGCATTTTTGAAAAATTCTGTTTTTATAATTTGCAAGTGAATATTTGGAGCACTTTGAAGCCTATGGTGGATAAGGAAATATTTCACATAAAAACTACACAGAAGCATTC
n33 169 CCCAGAAACTTTTTGTGATGTGTGCATTCATATCACATATTTGAACCTATCTTTTGATAGAGCAGTTTTGAAACTCTCTTTTTGTAGAATCTGGATATTTGGAGCCCTTTGTAGTCTATGGTGGAAAAGGAAATATCCTCACATAAAAACCACACAGAAGCATT
n34 171 TGAGAAAATTCTTTGTGATGTGTGCATTCAACTGACAGAGTTGAACCTTTCTTTTGATTGACTAGTTTTGAAAATCTCTTTTTGTAGAATCTGCAAGTGGATATTTGGAGTGCTTTGAGGCCTATGGTGGAAAAGGAAATATCTTCATATGAAAACTACACAGAAGCATTC
n35 170 CGAGAAACTTGTTTGTGATATGTGCATTCAGCTCACAGACTTGAAACTTCCTCTTGATTGAGCAGTTTGGAAACACTCTTTAGTAAAATCTGCAAGTGGATATTCGGAGCACTTTGAGGCCTGTTGTGGAGAAGGAAATATCTTCACATAAAAACTACACAGACGCATTC
n36 171 TCAGAAACTACTTTGTGATGTGTGCGTTCATATCACATAGTTGAACCTATCTTTTGATAGAGCACTTTTGAAACTCTCTTTTTGTAGAATCTGCAAGTGGATATTTGGAGCCCTTTGCAGCCTATGGTGGAAAAGGAAACATCTTCACATAAAAACTACACAGAAGCATTC
n37 171 TGAGAAACTACTTTGTGATGTGTGCATTCAACTGACAGATTTGAACCTTTCTTTTGATTAAATAGTTTTGAAAATCTCTTTTTGTAGAATCTGCAAGTGGATATTTGGAGTGCTTTGAGGCCTATGGTGGAAAAGGAAATATCTTTACATAAAAACTACACAGAAGCATTC
n38 171 CAATAAACTTGTTTGTGATATGTACCTTCACCTCACAGACTTGAAACTTCCTCTTGATTGAGCAGCTTGGAAACACACTTTTAGTGAAATCTGCAAGTGGATATTTGGAGCACCTTGAGGCCTGTTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGCATTC
n39 171 TCAGTAACTTCTTTGTAATGTGTGCATTCAACTCGCAGAGTTGAACCTATCTTTTGATACAGCAGTTTTGAAACTCTCTTTTTGTAGAATCTGCAAGTGGATATTTGCAGCGCTTTGAGGCCTGCGGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGCATTC
n40 169 CCAGAAACTTATTTGTGATGTGTTTATTCATCTCACGGAGTTGAACCTTTCTTTGATTGACAAGTTTTGAAAGACTATGTTTCTCTATAATGTGCAAGTGGATATTTGGAGTGCTTTGAGGCCATATGGTGGAAAAGGAAATATATTCACATAAAACTATACAGAAGCGTTC
n41 171 CGAGAAACTTCTCTGTGATGTGTGCACTAATCTCACAGAGTTTAATCTATCATTTGATTGAGCAGTTTTAAAAAACTTTTTTTGTGGAATCTGCAATTGGATATTTGGAACGCTTTGAGGCCTATTGTGGAAAAGGCAATATCTTCACATAAAAACTACACAGAAACATTC
n42 171 TGAGAAACTTTTTTGACATGTGTGCACTAAACTCTCAGAGTTGAACTTATCTTTTGATAGAGCAGTTTTGAAACTCTGTGTTACTAGAATCTGCATGTGGTTATTTGGAGTCCTTTGTGGCCGATGGTGGAAAAGGAAATATCTTCCCCTAAAAAGTACACAGAAGCATTC
n43 170 TGAGAAACTTCTTCACGATGGTTGCACTCATACCCACAGAGTCGAAACTATCGTTTGAGAGAGCATTTCGAAACTTTCTTTTTGTAGGATCTGCAAGTGGATATTTGGAGGGCTTTCAGGCCTATGGTGGAAAAGGAAATATCTTCACATAAACACTACTCAGAAGCATTC
n44 169 TGAGAAACTTCTTTGTGATGTGTGTATTTGTCTCAGACTGGAACCTTCCTTTTGATTGAGCAGTTCTGAAACACTCTTTTTGTAGAATCTGGAAGTGCATATTTGGAGTGCTTTGAGGCCTATGGTGGAAAAAGAAATATCTTCATTTAAAAACTACACAGAAGCATTC
n45 166 TGAGAAACTACCTTTTGATGTGTGTATTTCACAGCGTTAAACTTACCTTTGGTAGAGCAGTTTTGAAACTCTCTTTTTGTGGAAAATGTAAGTGGGTATTTAGAGCCATTTGTGGCCTATGGTGGAAAGGAAAATATCTTCACATAAAAACTACACAGAAGCATTC
```

Figure 1. Continued.

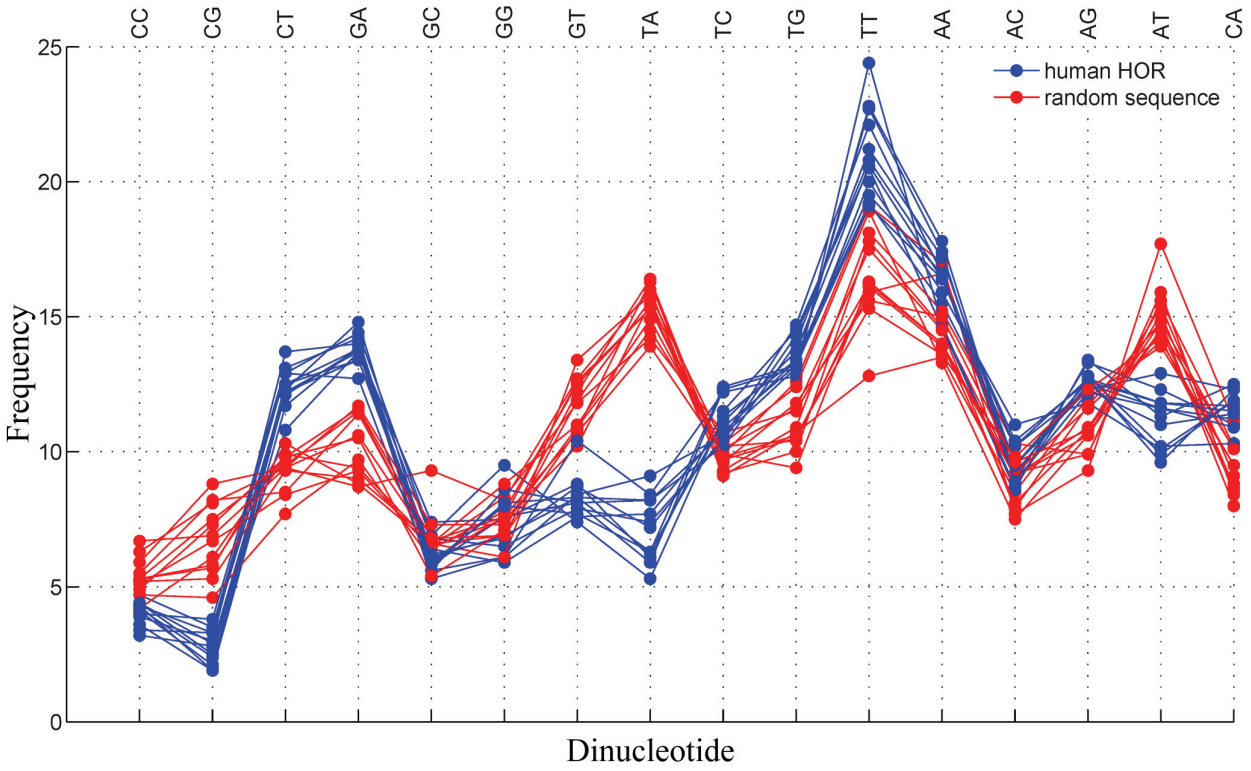

Figure 2. Dinucleotide frequencies in human consensus HOR alpha satellites and comparison to random sequences. Frequency of dinucleotides are shown for HOR alpha satellites (red) and for the corresponding random sequences (blue). It is seen that the TA-low effect is strongly pronounced in alpha satellite HORs.

Figure 3. Specific composition of start/stop CLTs in alpha satellite monomers from HOR and monomeric sequences in human chromosomes and in chimpanzee Y chromosome. Chromosomes without specifications of species are human, while chimpanzee chromosome is denoted as chimp. AS denotes alpha satellite.

Shannon construction of trinucleotide frequencies alpha satellite HORs is found here, they do not exhibit any periodic structure to sides of the central motif (Supplementary figure 2). Here an alternative model of wrapping around nucleosome is proposed, based on a two-fold role of start/stop CLTs, both as a component of regulatory element and as construction element of nucleosome structure.

In the framework of this approach the structure of alpha satellite monomer is organized in three segments (Figure 4): the first (~84 bp) and the second (~58 bp) wrapping around the nucleosome and a linker (~29 bp). On the basis of start/stop CLTs it is recognized that the start of the first wrapping is the first ATG start-CLT in the sequence of 171-bp alpha satellite monomer, and the start of the second wrapping is the second ATG CLT (in extended form ATTTGG) positioned precisely at the half of alpha satellite monomer. The first and the second wrapping sequence end with the stop-TAG CLT. After the second wrapping sequence follows the linker of ~29 bp, which ends with TG nucleotide. This dinucleotide is combined with ATG as the start of the next alpha satellite monomer. In this way the string TGATG is formed. As already mentioned, the TGATG represents stop-TGA and start ATG CLT with A-nucleotide overlap. It is demonstrated that the first 27 bases in the linker do not contain any start/stop CLT, what significantly distinguishes linker from the two wrapping sequences in alpha satellite monomers.

There is a question why so much effort should have been spent during evolution exclusively for building elements such as sophisticated chromosome specific HORs with highly specific constituent alpha satellites. Previously we suggested that this HOR specificity and form of DNA folding is necessary that particular microtubules be bound to centromere in the corresponding chromosomes.[31] Of course, the centromere should be substantially better organized and stronger so that it does not break to endanger very complex organisms such as primates and in particular *Homo sapiens* as the last and most sophisticated in the evolutionary chain. One might ask whether it is needed to use such a complex structure as HORs, instead of having simple microsatellite tandems without any chromosome specificity. Therefore, it could be argued that HORs may also play some regulatory roles in the network of genes and gene regulators.

Furthermore, the central position of alpha satellite HORs in chromosomes, deeply protected within specific structure of centromeres, points to their possible importance and potential influence on the whole DNA. This might be compared, in principle, with a biological organization present on macro-plan, for example, by pituitary gland as master gland, which is strongly protected in the "armour" of *sella turcica* in sphenoid bone of scull and has a dominant influence on all endocrine glands in the body at significant distances. CLTs having different lengths of extension give, among others, possibilities of significantly higher number of combinations and have sizably higher information potential, contributing to higher interspecies variability. In general one might argue that the presence of internal extended cluster organization of alpha satellites could enable detailed analysis and holistic approach for the whole genome.

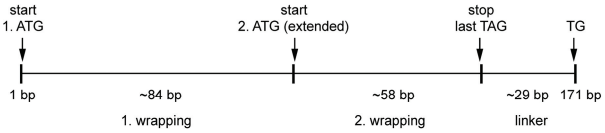

Figure 4. Start/stop CLTs scheme of wrapping alpha satellite around nucleosome and linker.

CONCLUSION

Extended start/stop codon-like clusters for noncoding sequences are introduced as a "new language" of DNA with rich information potential and shown to be most pronounced in alpha satellite higher order repeats, with a possible role in gene regulators.

This work shows in alpha satellite HORs an extended cluster organization based on T and/or A nucletide expansion in four dominant groups of trinucleotides which in coding parts of genomes represent start and stop codons. Such extended clusters are referred to as start/stop codon-like trinucleotides (CLTs). It is shown that what significantly distinguishes alpha satellite sequences in human centromere and pericentromere regions from non-alpha satellite HORs and from no-repeat noncoding sequences are the specificity and the level of extensions of start/stop CLTs, with dominant contribution from the stop-TGA CLTs. Successive multiplications of T nucleotides in stop-TGA CLTs, and to somewhat smaller extent due to multiplication of A nucleotides, leads to the poly-T class of all human alpha satellite HORs while the monomeric alpha satellites can be of poly-T or of poly-A class. The basic role of stop-TGA CLT is that the poly-T stretch preceding G nucleotide and the poly-A stretch following it (poly-T−G−poly-A) exhaust about 30 % of all T and A nucleotides in alpha satellites. Distributions of extended and non-extended CLTs within alpha satellites do not appear randomly, but form a well organized structure. In this way, this work opens a question of information potential of the new "CLT-language".

The CLT framework can provide also a structural function as nucleosome positioning patterns, giving a possible explanation of alpha satellite segmentation into two wrapping sequences around the nucleosome and a linker.

It is a question now what could be in centromeres a biological function of so sophisticated CLT structure, resembling to a new language of DNA with an additional information potential, because if there is a goal only to achieve the strength of centromere, this purpose could be satisfied in simpler ways. It could be hypothesized that CLTs in such a well organized form as in alpha satellites may be the related to regulative role in the genome, with noncoding RNAs transmitting the information signal to the gene network. This intriguing point might be of interest for future experimental investigations.

Supplementary Materials. – Supporting informations to the paper (Figures S1 and S2; Tables S1 to S3) are enclosed to the electronic version of the article. These data can be found on the website of *Croatica Chemica Acta* (http://public.carnet.hr/ccacaa).

Acknowledgements. The authors thank Chris Tyler-Smith for suggestion to look for possible new classifications of alpha satellites, and to anonymous reviewer for very useful comments.

REFERENCES

1. F. Jacob and J. Monod, *J. Mol. Biol.* **3** (1961) 318–356.
2. R. J. Britten and E. H. Davidson, *Science* **165** (1969) 349–357.
3. M. C. King and A. C. Wilson, *Science* **188** (1975) 107–116.
4. L. A. Pennacchio, N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Fick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, and E. M. Rubin, *Nature* **444** (2006) 499–502.
5. G. A. Wray and C. C. Babbitt, *Science* **321** (2008) 1300–1301.
6. D. A. Garfield and G. A. Wray, *BioScience* **60** (2010) 15–23.
7. J. P. Noonan and A. S. McCallion, *Annu. Rev. Genomics Hum. Genet.* **11** (2010) 1–23.
8. E. N. Trifonov, *Comput. Chem.* **17** (1993) 27–31.
9. L. Manuelidis and J. C. Wu, *Nature* **276** (1978) 92–94.
10. H. F. Willard, *Amer. J. Hum. Genet.* **37** (1985) 524–532.
11. C. Tyler-Smith, *Development* **101** (1985) 93–100.
12. C. Tyler-Smith and W. R. A. Brown, *J. Mol. Biol.* **195** (1987) 457–470.
13. P. E. Warburton and H. F. Willard, *Evolution of centromeric alpha satellite DNA: molecular organization within and between human and primate chromosomes*, in: M. Jackson, T. Strachan, and G. Dover (Eds.), *Human Genome Evolution,* BIOS Scientific, Oxford, 1996.
14. M. K. Rudd, G. A. Wray, and H. F. Willard, *Genome Res.* **16** (2006) 88–96.
15. R. Nussinov, *Nucleic Acids Res.* **12** (1984) 1749–1763.
16. E. Beutler, T. Gelbart, J. Han, J. A. Koziol, and B. Beutler, *Proc. Natl. Acad. Sci. USA* **86** (1989) 192–196.
17. S. Karlin and J. Mrazek, *Proc. Natl. Acad. Sci. USA* **94** (1997) 10227–10232.
18. S. Karlin, *Curr. Opin. Microbiol.* **1** (1998) 598–610.
19. P. F. Arndt, C. B. Burge, and T. Hwa, *J. Comput. Biol.* **10** (2003) 313–322.
20. C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, *Nature* **356** (1992) 168–170.
21. Y. Almirantis and A. Provata, *Bull. Math. Biol.* **59** (1997) 975–992.
22. E. Segal and J. Widom, *Curr. Opin. Struct. Biol.* **19** (2009) 65–71.
23. C. K. Collings, A. G. Fernandez, C. G. Pitschka, T. B. Hawkins, and J. N. Anderson, *PLoS One* **5(6)** (2010) e10933.
24. G. Levinson, and G. A. Gutman, *Mol. Biol. Evol.* **4** (1987) 203–221.
25. M. Rosandić, V. Paar, and I. Basar, *J. Theor. Biol.* **221** (2003) 29–37.
26. M. Rosandić, V. Paar, M. Glunčić, I. Basar, N. Pavin, *Croat. Med. J.* **44** (2003) 386–403.
27. V. Paar, N. Pavin, I. Basar, M. Rosandić, I. Luketin, and S. Durajlija Žinić, *Croat. Chem. Acta* **77** (2004) 73–81.
28. V. Paar, N. Pavin, M. Rosandić, M. Glunčić, I. Basar, R. Pezer, and S. Durajlija Žinić, *Bioinformatics* **21** (2005) 846–852.
29. M. Rosandić, V. Paar, I. Basar, M. Glunčić, N. Pavin, and I. Pilaš, *Chromosome Res.* **14** (2006) 735–753.
30. V. Paar, I. Basar, M. Rosandić, and M. Glunčić, *Curr. Genomics* **8** (2007) 93–111.
31. M. Rosandić, M. Glunčić, V. Paar, and I. Basar, *J. Theor. Biol.* **254** (2008) 555–560.
32. V. Paar, M. Glunčić, I. Basar, M. Rosandić, P. Paar, and M. Cvitković, *J. Mol. Evol.* **72** (2011) 34–55.
33. V. Paar, M. Glunčić, M. Rosandić, I. Basar, and I. Vlahović, *Mol. Biol. Evol.* **28** (2011) 1877–1892.
34. A. E. Rapaport, Z. M. Frenkel, and E. N. Trifonov, *J Biomol Structure & Dynamics* **4** (2011) 567–574.
35. Z. M. Frenkel, T. Bettecken, and E. N. Trifonov, *BMC Genomics* **12** (2011) 203.