

Human alpha satellite DNA – consensus sequence and conserved regions

B.Vissel and K.H.Choo

Murdoch Institute, Royal Children's Hospital, Flemington Road, Parkville 3052, Australia
Submitted June 29, 1987

Since the previous human alpha satellite concensus sequence (1) was derived by direct sequencing of genomic DNA containing the EcoRI-dimer structure, the other major alpha satellite species were naturally excluded. We report a modified human alpha satellite DNA concensus sequence (Table 1) derived from published data for 145 individually cloned and sequenced, complete alphoid monomers: 4 monomers from chromosomes 13 and 21 (2), 2 from acrocentric chromosomes (3), 3 from chromosome 22 (4), 5 from the Y chromosome (5), 16 from chromosome 17 (6), 12 from the X chromosome (7), 32 from chromosome 7 (8) and 71 from the 340 bp band in EcoRI-digested genomic DNA (8). Since these monomeric units were derived from a range of chromosomal sources, the new concensus sequence (Table 1) should be an improved representative of the human alpha satellite DNA (Table 2). Direct comparison of individual sequence also allows visualisation of nucleotides or regions of nucleotides which are well conserved (Table 1). We propose that the present concensus sequence should serve as a base for future updating as more sequence data becomes available.

Table 1 : Derivation of new concensus sequence

10														20														30													
I	A	A	T	T	C	T	C	A	G	T	A	A	C	T	T	C	C	T	T	G	T	G	T	T	G	T	G	T	G	T	30										
II	52	141	135	139	140	131	131	143	134	42	137	133	92	136	98	134	21	134	135	138	139	141	42	142	140	124	106	135	134	36											
III	93	4	10	6	5	14	14	2	11	103	8	12	53	9	47	11	124	11	10	7	6	4	103	3	5	21	39	10	11	109											
IV	76C	*	*	*	*	*	*	*	*	101A	*	*	39A	*	*	119T	*	*	*	*	*	*	*	102A	*	*	*	*	*	*	104C										
V	(C)	A	T	T	C	T	C	A	G	(A)	A	A	C	T	T	C	(T)	T	T	G	T	G	(A)	T	G	T	G	T	G	(C)											
40														50														60													
I	A	T	T	C	A	A	C	T	C	A	C	A	G	A	G	T	T	G	A	A	C	G	A	T	C	C	C	T	T	T	A										
II	86	142	144	136	140	130	128	144	142	132	137	133	125	143	144	136	116	87	138	142	131	36	12	119	73	139	132	133	141	46											
III	59	3	1	9	5	15	17	1	3	13	8	12	20	2	1	9	29	58	7	3	14	109	133	26	72	6	13	12	4	99											
IV	47C	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	51T	*	*	73C	127T	63T	*	*	*	*	*	*	*	85T											
V	A	T	T	C	A	A	C	T	C	A	C	A	G	A	G	T	T	G	A	A	C	(C)	T	T	C	C	T	T	T	T											
70														80														90													
I	C	A	C	A	G	A	G	C	A	G	A	C	T	T	G	A	A	A	C	A	C	T	C	T	T	T	T	T	G	T											
II	71	137	46	126	142	134	135	142	139	141	38	12	123	66	142	141	142	143	133	133	132	133	139	136	90	142	134	140	139	111											
III	74	8	99	19	3	11	10	3	6	4	107	133	22	79	3	4	3	2	12	12	13	12	6	9	55	3	11	5	6	34											
IV	71G	*	84T	*	*	*	*	*	*	91T	130T	59G	*	*	*	*	*	*	*	*	*	*	*	*	*	50G	*	*	*	*	*	*									
V	C	A	(T)	A	G	A	G	C	A	G	(T)	(T)	T	T	G	A	A	A	C	A	C	T	C	T	T	T	T	T	G	T											
100														110														120													
I	G	G	A	A	T	T	T	G	C	A	A	G	T	G	G	A	G	A	T	T	T	C	A	G	C	C	G	C	T	T											
II	49	90	137	91	141	50	141	136	137	143	141	135	135	139	132	138	45	141	135	118	123	43	26	15	83	120	71	113	119	135											
III	96	55	9	54	4	95	4	9	8	2	4	10	10	6	13	7	100	4	10	27	22	102	119	130	62	25	74	32	26	10											
IV	91A	50A	*	49G	*	91C	*	*	*	*	*	*	*	*	*	*	*	*	90T	*	*	72G	98G	129A	55G	45T	*	*	*	*	*	*									
V	(A)	G	A	A	T	(C)	T	G	C	A	A	G	T	G	G	A	(T)	A	T	T	T	(G)	(G)	(A)	C	C	G	C	T	T											

I	T	G	A	G	G	T	C	A	A	T	G	G	T	A	G	A	A	T	A	C	G	A	A	A	T	C	T	T		
II	141	133	126	109	138	29	120	43	59	44	108	60	140	47	143	140	143	8	46	142	140	65	135	72	131	84	134	127	138	140
III	4	12	19	36	7	116	25	102	86	101	37	85	5	98	2	5	2	137	99	3	5	80	10	73	14	61	11	18	7	5
IV	*	*	*	*	97C	98T	78T	94C			79T	*	89G	*	*	130A	78C	*	*	74G	66T	59-	*	*	*	*				
V	T	G	A	G	G	(C)	C	T	A	(C)	G	G	T	(G)	G	A	A	(A)	(C)	G	G	A	A	T	C	T	T			
I	C	C	T	A	T	A	G	A	A	A	C	T	A	G	A	C	A	G	A	A	T									
II	142	14	82	103	134	91	8	82	78	63	136	106	138	115	145	135	138	140	141	142	39									
III	3	131	63	42	11	54	137	63	67	82	9	39	7	30	0	10	7	5	4	3	106									
IV	*	102A	40C	29T	*	120A	47T	49G	63-	*	*	*	*	*	*	*	*	*	*	*	95G									
V	C	(A)	T	A	T	A	(A)	A	A	A	C	T	A	G	A	C	A	G	A	A	(G)									

- I - Monomer-1 of the EcoRI dimer concensus (1).
 II - No. of bases in 145 alploid monomers that are identical to the concensus sequence in I.
 III - No. of bases in 145 alploid monomers that are different from the concensus sequence in I.
 IV - Most common alternative to a given concensus base in I (when present in greater than 25% of monomers).
 (*) indicates concensus base is conserved in greater than 90% of monomers.
 V - Derived concensus sequence. A base present in greater than 50%, or two bases each present in greater than 40% of monomers, is regarded as concensus. Bases in brackets represent modification from previous concensus.

Table 2. Comparison of concensus sequences with different alpha satellite monomers.

	R1	R2	X	Y1	17/1	7-1/1	7-1/2	L1.26/1	S5/1	AGM
CON 1	85	91	94	84	84	84	92	88	91	73
R1		73	78	73	71	85	73	76	77	67
R2			84	73	77	73	91	76	78	65

The values represent percent homology. CON1 = derived concensus (Table 1); R1, R2 = EcoRI dimer concensus (1); X = X-chromosome concensus (7); Y1 = a monomer from Y-chromosome (5); 17/1 = monomer 1 of p17H8 from chromosome 17 (6); 7-1/1, 7-1/2 = monomers 1 and 2, respectively, of EcoRI (340 bp) dimer from chromosome 7 (8); L1.26/1 = a monomer from chromosomes 13 and 21 (2); S5/1 = a monomer from acrocentric chromosomes (3); AGM = African Green Monkey concensus sequence (9). Note the general improvement of sequence homology with CON1 compared to previous concensus R1 and R2.

References

- WU J. and MANUELIDIS L. (1980) J. Mol. Biol. 142, 363-386.
- DEVILEE P. et al. (1986) Nucl. Acids Res. 14, 2059-2073.
- CHOO K H. et al. (in preparation).
- McDERMID M E. et al. (1986) Chromosoma (Berl.) 94, 228-234.
- WOLFE J. et al. (1985) J. Mol. Biol. 182, 477-485.
- WAYE J S. and WILLARD H F. (1986) Mol. Cell Biol. 6, 3156-3165.
- WAYE J S. and WILLARD H F. (1985) Nucl. Acids Res. 13, 2731-2743.
- JORGENSON A L. et al. (1986) J. Mol. Biol. 187, 185-196.
- ROSENBERG H. et al. (1978) Science 200, 394-402.