JCBPS; Section B; August 2017 – October 2017, Vol. 7, No. 4; 1011-1036. E- ISSN: 2249 –1929

[DOI: 10.24214/jcbps.B.7.4.101136]

Journal of Chemical, Biological and Physical Sciences

An International Peer Review E-3 Journal of Sciences

*Available online at*www.jcbsc.org

Section B: Biological Sciences

CODEN (USA): JCBPAT Research Article

Few Simple Sequence Repeats in Human Hair Keratin Genes

Jitendra Gharu, Seema Trivedi *

Department of Zoology, JN Vyas University, Jodhpur (Rajasthan), India

Received: 13 July 2017**; Revised:** 27 July 2017**; Accepted:** 08 August 2017

Abstract: Keratin genes are subgroup of intermediate filament (IF) genes. Keratins expressed in hair are called "hair keratins". Comparison of keratin gene sequences and structural organization has been done earlier but no study has been undertaken that compares human hair keratin genes with orthologues in chimp, orangutan, macaque and platypus, with specific references to simple sequence repeats (SSRs) or microsatellites. This study seeks presence of SSRs in human hair keratin genes and compares their positions in orthologue genes. Although the structural organization of keratin genes is largely conserved, as reported in other studies; some human keratin genes show differences in the number of exons and the lengths of exons/introns, when compared with the orthologues. The present study shows few SSRs in human hair keratin genes, where only one repeat is found in exon of Type I keratin gene KRT31, while additional repeats are present in introns. Some repeats are longer in the human genes whereas some indicate reduction in length compared with orthologues. Many repeats are present in regions that are known for intronic variations. This study revisits the comparisons of structural organization of human hair keratin genes and compares them with the orthologues. The SSRs found here could be useful for monitoring variations in human hair keratin genes that may happen due to hypermutable nature of SSRs.

Keywords: Exon Length, Gene Structure, Hair Keratin Genes, Intron, Microsatellite(s), Simple sequence repeat(s) (SSRs), Repeat Conservation.

INTRODUCTION

Keratin genes and proteins have been identified as a subgroup of intermediate filaments (IF) genes and proteins [1]. The keratins expressed in hair are called "hard", "trichocytic" or "hair" keratins. These differ from other keratins in having higher amount of sulphur in the head and tail domains [2, 3] and cysteine residue compared with epithelial keratins [3-6]. In humans there are seventeen hair keratin genes encompassing eleven Type I keratins and six Type II keratins [5-9]. Though keratin genes of the same type are closely linked, Type I and Type II are present on human chromosome 17q12-21 and 12q13, respectively [1, 2]. Keratin genes have similarity in sequences, exon/intron structure [8] and conserved elements in promoter region [10]. Type I hair keratin genes have six introns except hHa5 (KRT35) which has a seventh intron in the 3' noncoding region [11, 12]. Type II hair keratin genes have nine exons and eight introns; the exon-intron boundaries are conserved. The exception is hHb4 gene with nine introns which is in the 3' noncoding region [12]. Moreover, most introns in Type II keratin genes begin with 5' GT and end with 3' AG nucleotide combination [13]. Many disorders are associated with mutations in human keratin genes [14-17]. Mutations in Type II hair keratins that result in disorders are reported but none so far in Type I keratins [2, 4, 9, 18]. Perusal of literature indicates that presence of simple sequence repeats (SSRs) in hair keratin genes in humans has not been investigated. Repeated units of DNA nucleotides called SSRs or microsatellites are present in many prokaryotic and eukaryotic genomes. These repeated units may consists of motifs ranging from 1-10 nucleotides [19-24]. SSRs are considered hypermutable sequences where mutations could be due to insertions, deletions of a motif or of a single nucleotide in a motif [22, 25]. Expansion or contraction of SSRs generally happens during DNA replication and is frequent on the lagging strand [21, 26-29]. SSR length mutations may modify genes [21, 30, 31] and or cause disorders whether repeats are present in exons, introns or un-transcribed regions [23, 32-40]. Triplet repeat expansions can also lead to non-ATG (RAN) translation i.e. can initiate translation without ATG codon [41].

However, SSRs may have functional roles [23, 25, 30, 38, 42-45]. Selection of microsatellites can affect linked gene diversity [46], genome diversity [47], influence evolution [19, 30, 31, 38, 48], rate of evolution of sequences (gene) [49], adaptability to changing environmental conditions [25], brain development and function [38] etc. SSRs can also influence phenotype or behaviour at individual and interspecies level as seen in study on vasopressin *avpr1a* gene in voles (*Microtus*), homologue in humans, bonobos (*Pan paniscus*) and chimp (*Pan troglodytes*) [30]. SSRs may also influence facial phenotype as found in case of dog breeds [30, 50]. Similarly, polydactyly in a dog breed (Great Pyrenees) is associated with a hexanucleotide repeat reduction in *Alx-4* gene [30, 50]. The present study examines SSR type, length, CG/GC richness and positions in human hair keratin Type I and Type II genes. Further, we compare the positions of SSRs in human hair keratin genes with those of chimp, orangutan, macaque and platypus orthologue sequences (if present) to seek conservation of human SSRs. The SSRs found in the present study could be useful for further research for monitoring variations in human hair keratin genes that may happen due to hypermutable nature of SSRs. The present study also revisits comparison of structural organization of human hair keratin genes with those of orthologues in chimp, orangutan, macaque and platypus.

MATERIALS AND METHODS

Obtaining keratin gene IDs: Human (*Homo sapiens*) hair keratin gene HGNC (HUGO Gene Nomenclature Committee) IDs were obtained from NCBI database (https://www.ncbi.nlm.nih.gov/gene/). These IDs were submitted to Ensemble genome database version 75 [51, 52] (http://www.ensembl.org) to obtain Ensemble gene IDs and features like chromosome number, strand, GC%, gene length, transcript numbers etc. associated with each gene.

Obtaining gene sequences: Ensemble gene IDs of human hair keratin genes were used for obtaining unspliced gene sequences from Ensemble genome database 75 [51, 52].

Repeat search: Human hair keratin genes were processed for SSR search by using SciRoKo software [53]. Default parameters were used for repeat search, except for choosing mismatched fixed penalty mode and upper boundary of motif length was set at six. Search outputs were analysed on basis of motifs, lengths, CG richness and position in the gene sequences.

Repeat densities in gene sequences and exons/introns: Repeat density per KBp (Kilo Base pairs) in each gene was determined by dividing total number of repeats with gene length for each hair keratin gene. Presence of repeats in introns and exons was determined on basis of co-ordinates of exons for each hair keratin gene obtained from Ensembl 75 [51, 52]. Repeat density per KBp in exons and introns was determined by dividing total repeat numbers with total exon and intron lengths respectively for each gene.

SSR length and CG richness: Length of each SSR was measured based on number of nucleotides between repeat start and repeat end positions in each gene as per SciRoKo output. For SSR CG richness (percentage), total numbers of occurrences of C/G in each SSR unit were considered.

Conservation of repeats: Multiple aligned sequences [EPO (Enredo, Pecan, Ortheus)] of human hair keratin genes with orthologue gene sequences in chimp (*Pan troglodytes*), orangutan (*Pongo abelii*), macaque (*Macaca mulatta*) and Pecan aligned sequences of platypus (*Ornithorhynchus anatinus*) were obtained from Ensembl 75 [51, 52, 54, 55]. Human repeats and their positions were compared with the aligned sequences of orthologues. Repeats found in the orthologue sequences in the same position as those found in the human hair keratin gene sequences were considered conserved. Total number of such repeats from each orthologue sequence was taken for calculation of percentage of conserved repeats out of total human repeats. Similarly, repeat types and motif conservation percentage was calculated.

Statistical analysis and representation of data: Pearson's correlation analysis at 99% and 95% confidence was done with help of SPSS (statistical analysis software) to assess correlation between gene sequence length and CG% with repeat density, CG% and length of repeats. Heat maps were generated for representation of repeat density and conservation values by using Matrix2png [56].

Comparison of gene structure: Human hair keratin gene sequences aligned with orthologue sequences obtained from Ensembl version 75 were used for comparisons. Exons and introns in each human keratin gene sequence were compared with exons and introns in respective orthologue sequences from chimp, macaque, orangutan and platypus. These positions and conservation of sequences were confirmed from exon co-ordinates, exon ranks and the markings obtained from Ensembl 75 [51, 52, 54, 55]. Exon start and end co-ordinates and exon rank in transcripts obtained for each human hair keratin gene and respective orthologues were also used for calculating total number of exons for each gene and individual exon length, which were possible for each transcript variant for respective gene. Each exon length for every exon rank for respective gene was used for comparison of exon lengths of human genes with the orthologues. Equal to or less than 10nt differences in lengths of exons were ignored.

RESULTS AND OBSERVATIONS

Repeats in gene sequences: Out of 17 hair keratin genes, 63.64% of Type I genes and 83.33% of Type II genes have repeats. Four Type I keratin genes KRT33A, KRT35, KRT37 and KRT38 and one Type II (KRT84) gene sequences do not have repeats.

Repeat density is highest in KRT40 (0.8493/KBp) followed by KRT31 (0.5169/KBp) and KRT39 (0.4719/KBp). Among repeat types, KRT40 has highest density of di- and tetra-nucleotide repeats (0.3185/KBp) (**Figure 1**).

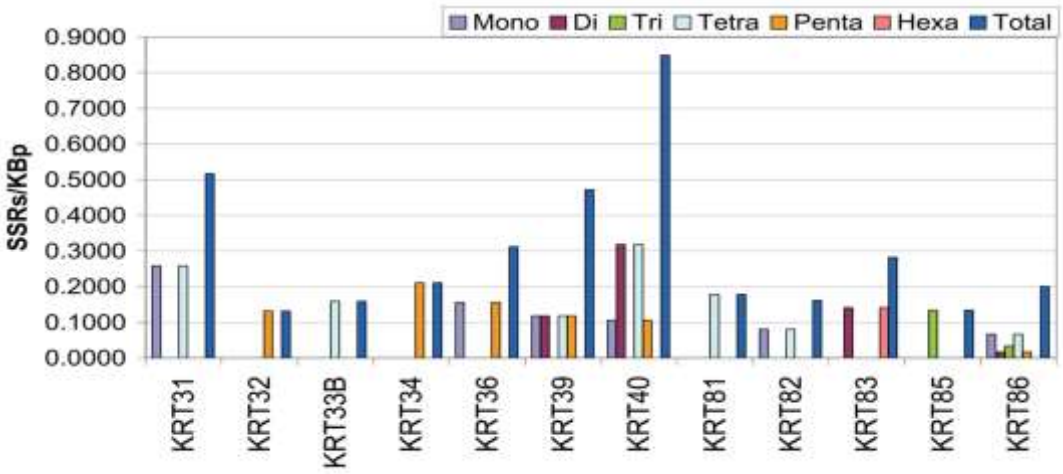

Fig.-1: Densities (per KBp) of total repeats and repeat types in human hair keratin genes

Amongst different motifs found in hair keratin genes, mononucleotide repeat T (n) and tetranucleotide repeat motif CTCC (n) (0.2585/KBp) have highest density present in KRT31 gene sequence (**Figure 2**). Pearson's correlation analysis reveals significant ($P < 0.01$) negative correlation between gene length and repeat density. Repeat GC richness shows significant ($P < 0.05$) positive correlation with repeat length.

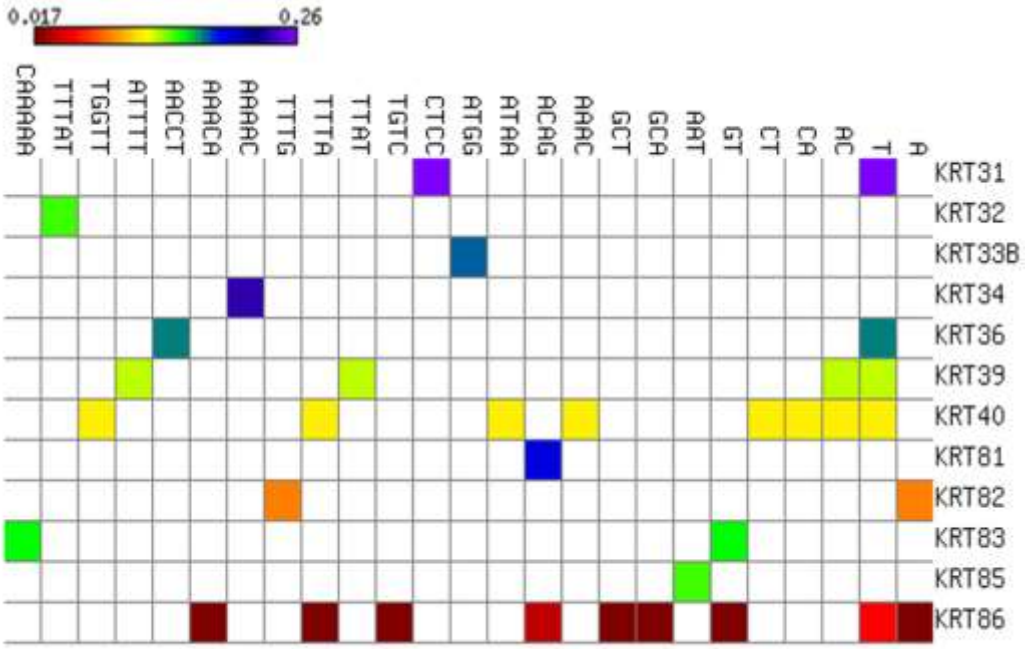

Fig.-2: Total motif densities (per KBp) in human hair keratin genes. Scale bar colour intensity indicates lower to higher values. Blank cells indicate absence of repeats and motifs

Repeats in exons and introns: Repeats are abundant in introns of hair keratin genes but present in exons of

only one gene i.e. KRT31 (tetranucleotide repeat). Repeat density is highest in introns of KRT40 (1.1117/KBp) followed by KRT39 (0.6115/KBp) (**Figure 3**). In introns, mononucleotide (Tn) repeats are the most abundant repeat types and present in KRT31 genes (0.4398/KBp). The other abundant repeat types in introns are di- and tetra-nucleotide repeats (0.4169/KBp respectively) present in KRT40 **Figure 4**).

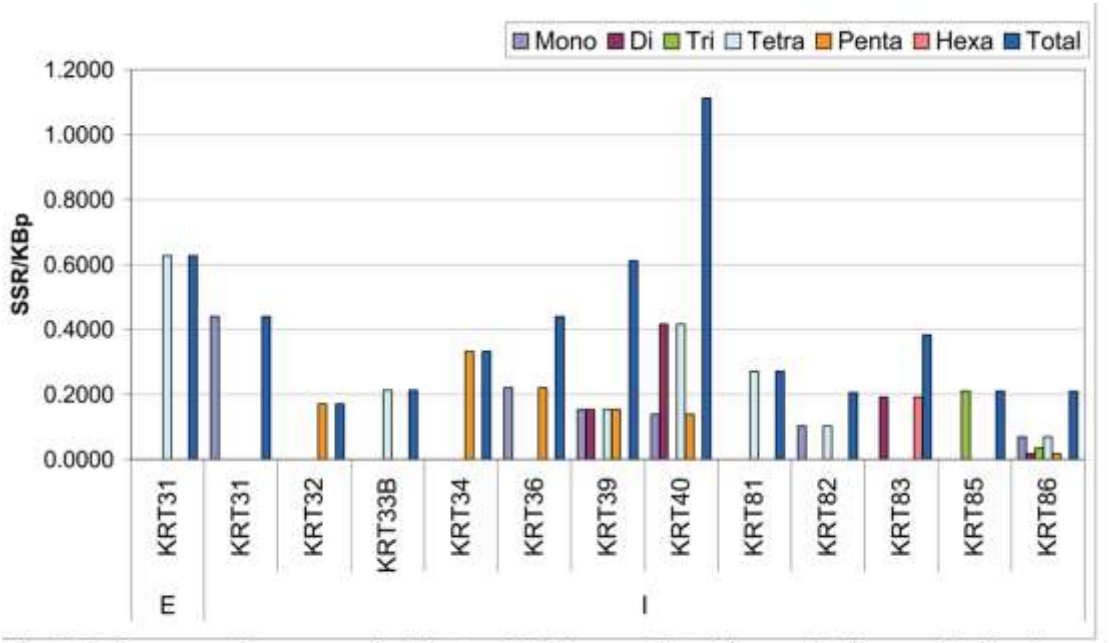

Fig.-3: Total repeats and repeat type densities (per KBp) in exons (E) and introns (I) of human hair keratin genes

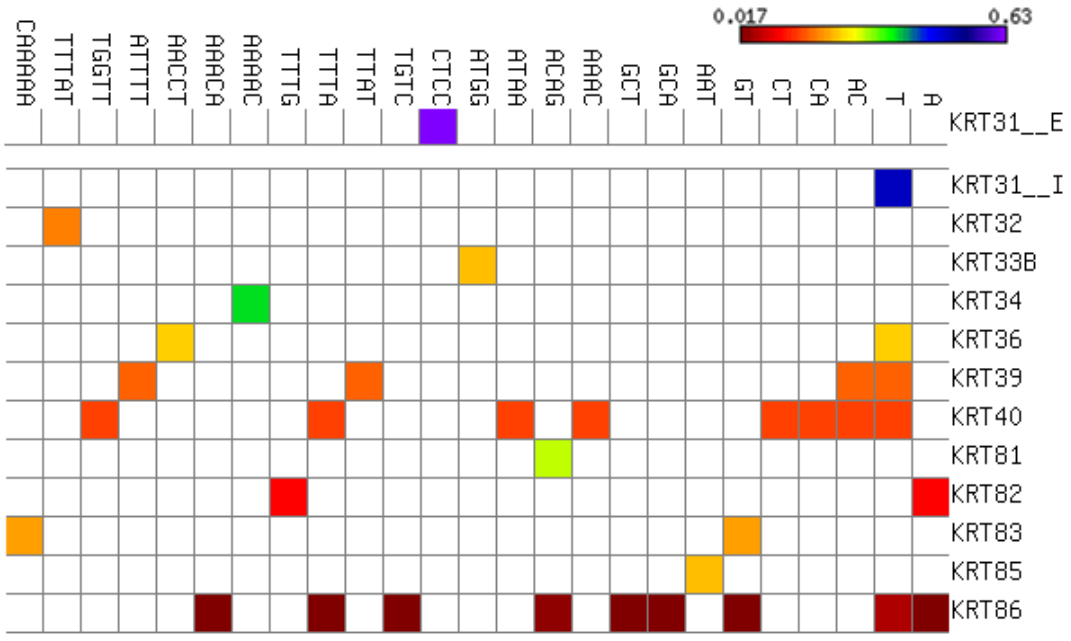

Fig.-4: Total motif densities (per KBp) in exons (E) and introns (I) of human hair keratin genes. Scale bar colour intensity indicates lower to higher values. Blank cells indicate absence of repeat and motif

Repeat length: Longest repeat in human hair keratin gene sequences is a pentanucleotide motif AACCT

(54nt) which is present in Type I KRT36 (**Figure 5**). This is followed by di- and tetra-nucleotide repeats in Type II hair keratin genes KRT83 and KRT86, which are 45nt and 36nt long GT and ACAG motifs respectively (**Figure 6**).

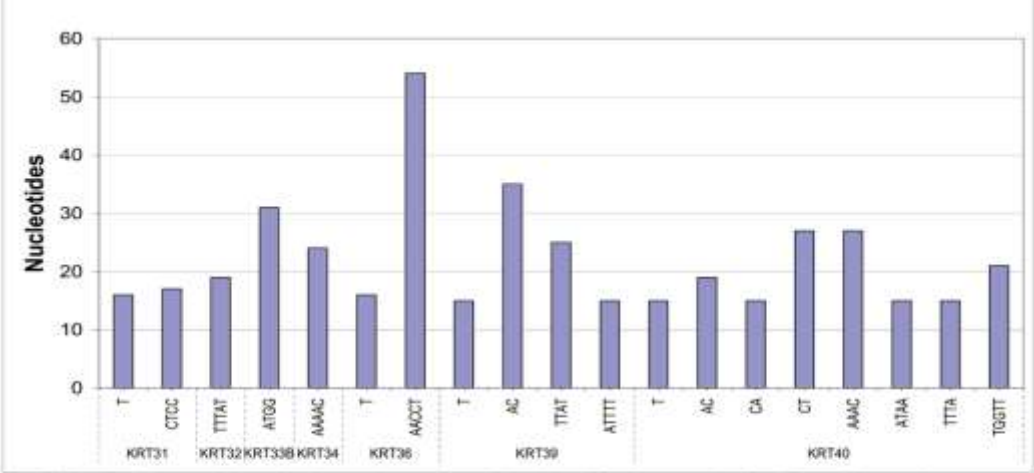

Fig.-5: Repeat lengths (number of nucleotides) in human hair Type I keratin genes.

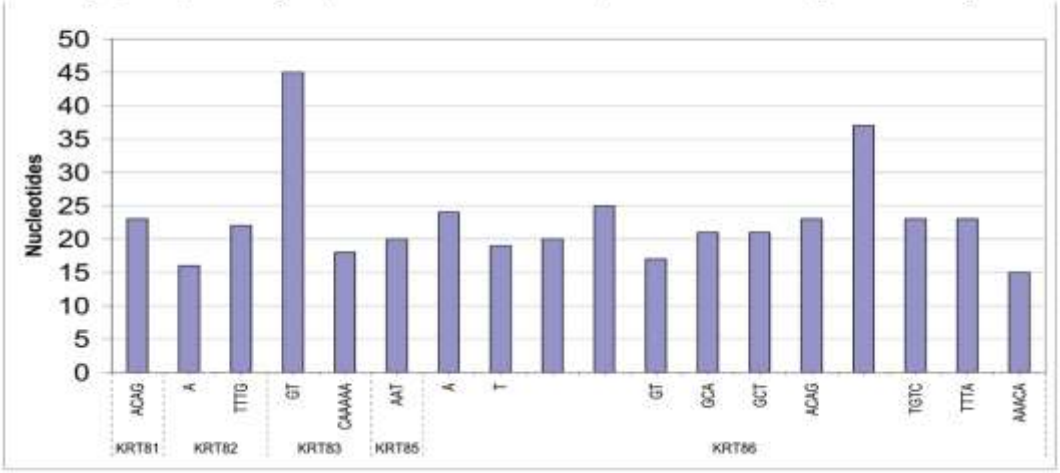

Fig.-6: Repeat lengths (number of nucleotides) in human Type II hair keratin genes.

CG richness of repeats: In general, AT rich repeats are prevalent in keratin gene sequences. Therefore, most repeats are not rich in CG content. Mononucleotide repeats and larger repeat motifs like hexa- and penta-nucleotide repeats are CG poor (**Figure** 7). KRT32 and KRT85 are unique in having no repeat motif with C/G nucleotide. Only two genes, namely KRT31 and KRT86 contain repeat motifs having >50% CG richness. KRT36 and KRT34 repeat motifs have 40% and 20% C/G richness respectively.

Repeat conservation: Total repeat conservation is lowest in KRT86. Only 8.33% of the total repeats are in the same position as those present in human sequences in orthologues of human KRT86 gene sequences in chimp, macaque and platypus (**Figure 8**). Conservations of repeat motifs in hair keratin genes show variations in orthologue sequences (**Figure 9**). However, repeat motifs CA, GCA, GCT, ATAA, TGTC, AAACA and TGGTT are not conserved in hair keratin orthologues. Thus, these appear to be human-specific repeat motifs.

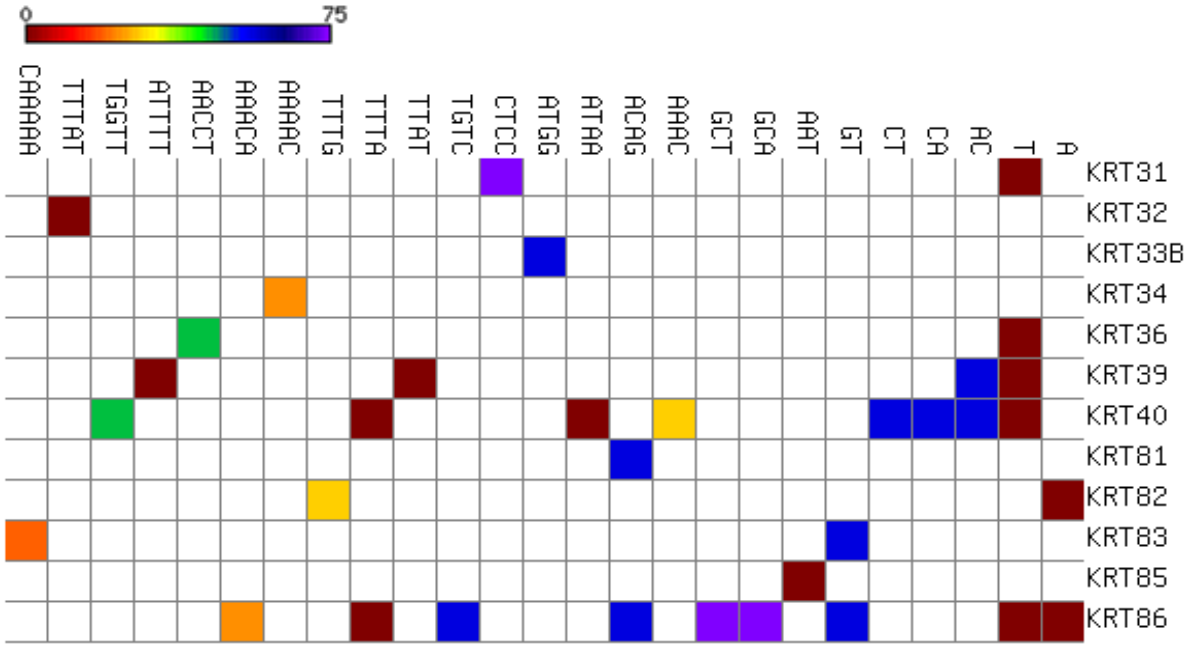

Fig.-7: Repeat CG% in human hair keratin genes. Scale bar colour intensity indicates lower to higher values. Blank cells indicate absence of repeat and motif

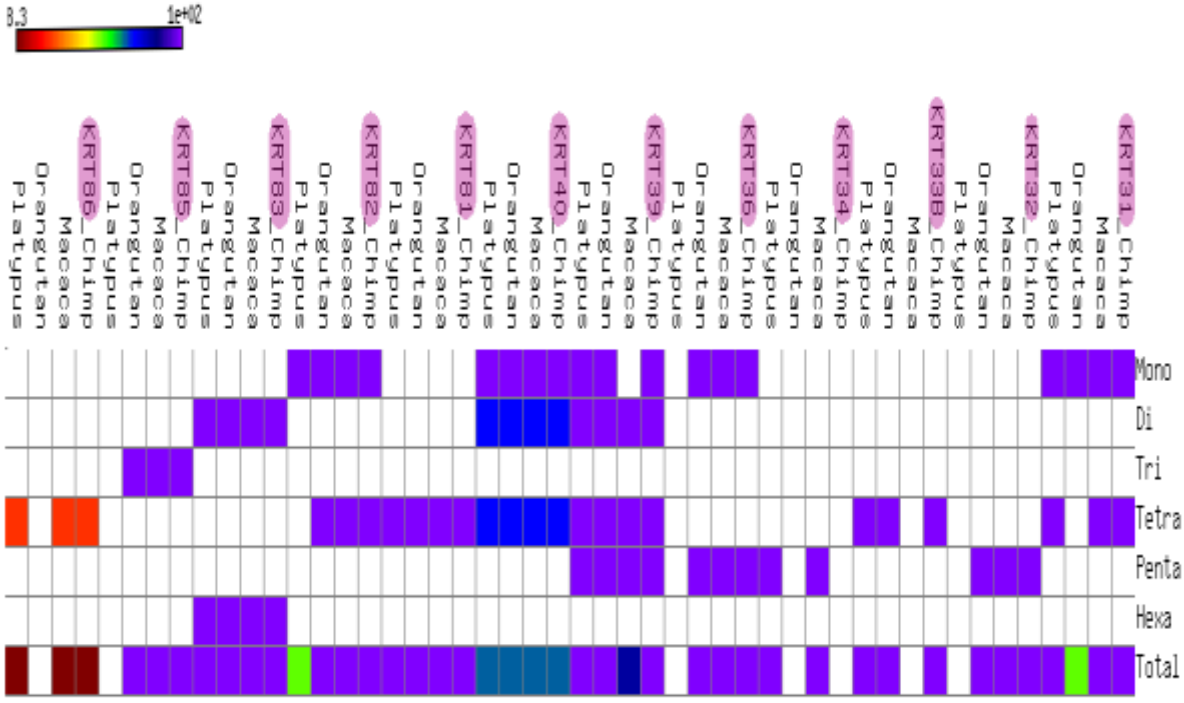

Fig.-8: Total repeat and repeat type conservation percentage of human hair keratin genes with reference to mammalian orthologue genes. Scale bar colour intensity indicates lower to higher values. Blank cells indicate absence of repeat and repeat type or zero value

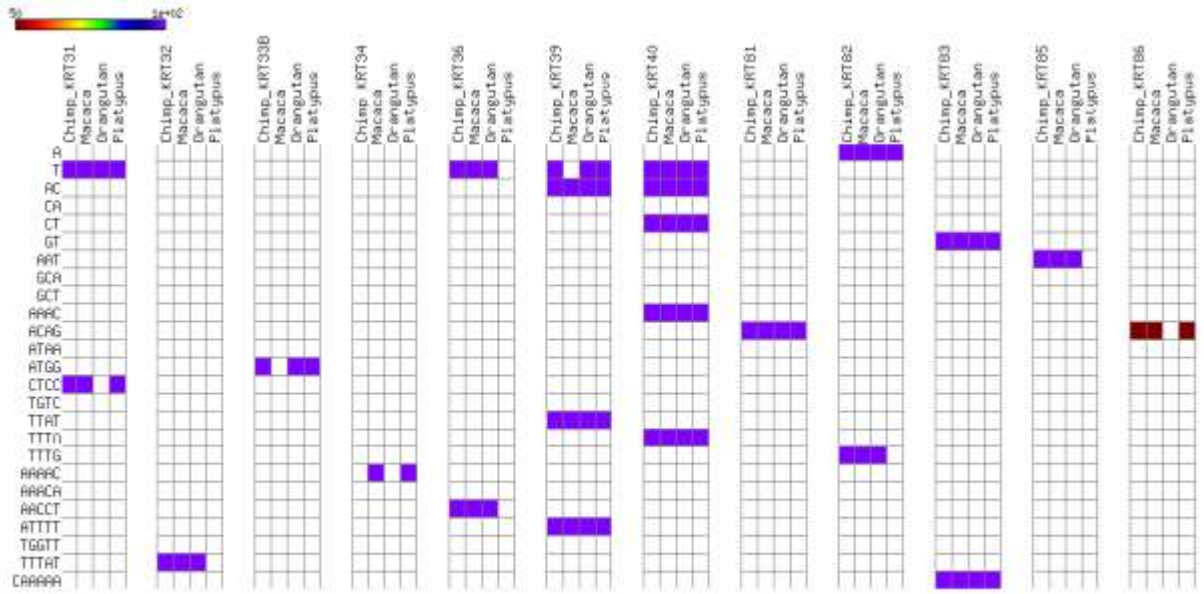

Fig.-9: Repeat motif conservation percentage in human hair keratin genes with reference to mammalian orthologue genes. Scale bar colour intensity indicates lower to higher values. Blank cells indicate absence of repeat and motif

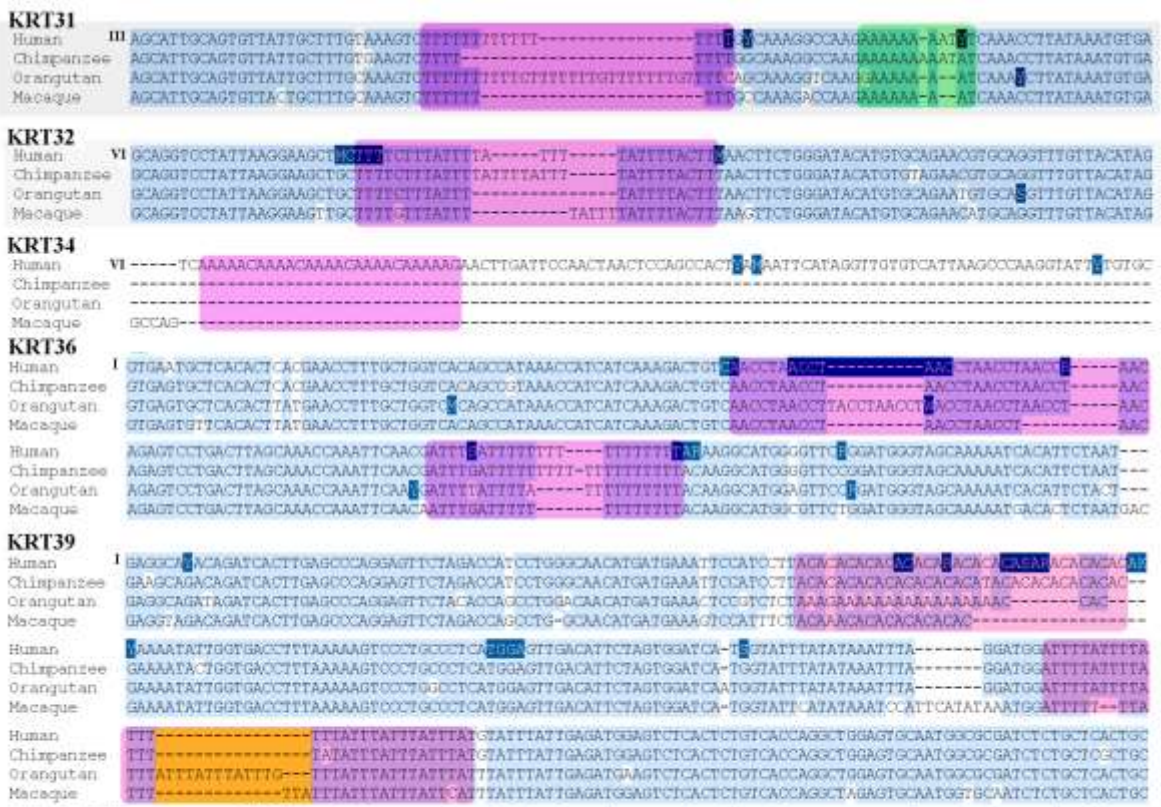

Fig.-10: SSRs (magenta highlight) in human Type I hair keratin genes in aligned sequences with orthologues in chimp, orangutan and macaque. Some sequences that are not SSRs but are repetitive (green highlight) are also seen. Longer repeat in chimp and orangutan orthologues are evident. Orangutan KRT39 intron I shows difference in motif of first SSRs. Longer repeat in KRT39 is marked with orange highlight. (Aligned sequences from Ensembl[51,52])

SSR comparison: Human KRT31 aligned with orthologues show longer intronic repeat (III intron) in

orangutan but shorter in chimp sequences (**Figure 10**). KRT32 (TTTAT)n repeat in intron VI is longer in chimpanzee sequence. Human KRT34 intron VI (including the SSR) has no sequence similarity with any orthologue. KRT36 pentanucleotide repeat (AACCT)n shows conservation in some regions but is shorter compared with orangutan, and (T)n repeat is shorter compared with chimp sequence. Except for mononucleotide repeat (T)n in intron I of human KRT39, other repeat types do not show conservation.

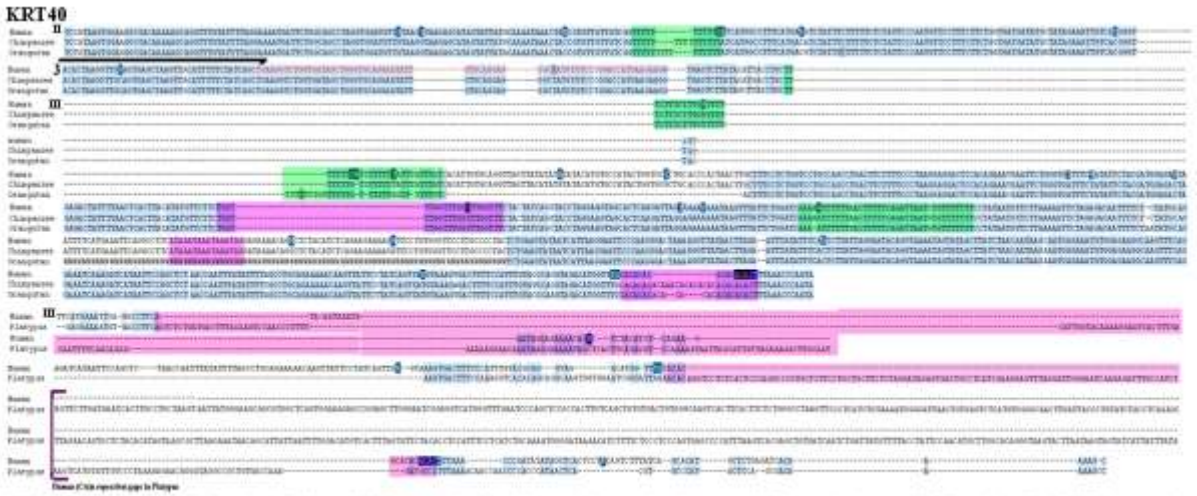

Fig.-11a: SSRs (magenta highlight) in intron iii of human KRT40 compared with orthologues and some regions that are not SSRs but repetitive (green highlight) in intron II and III. Human exon 3 (partial sequence, red font) does not show exon marking in the same region. Platypus homologue shows little similarity. (Aligned sequences from Ensembl[31,32])

Fig. 11b: SSRs (magenta highlight) and repeat like sequence regions (green highlight) in introns IV, V, VI and VIII of KRT40 aligned sequences. (Aligned sequences from Ensembl[31, 32]).

The regions that show similarity with human KRT40 exon 3 and introns II and III have human SSRs and

repeat like regions (**Figure 11a**). Further, all nucleotides of repeat (ATAA)n (magenta highlight) in intron III in human and chimp sequences show no differences, but the platypus sequence region corresponding to the same repeat has no similarity. The (CA)n repeat (magenta highlight) is short in human sequence compared with chimp sequence and orangutan repeat. The platypus sequence in this region has no similarity. Human KRT40 exon 5 and introns IV, V, VI and VIII have repeats (magenta highlight) and regions that could be repetitive in nature (green highlight) (**Figure 11b)**.

Comparisons show that (TTTA)n repeat near exon 5 and within intron V is longer in the macaque sequence. (CT)n repeat within intron V is longer in chimp and macaque genes and (T)n repeat within intron VIII is longer in orangutan and absent in macaque genes. However, the (AAAC)n repeat within intron VI is longer in the human gene (**Figure 11b)**. Type II keratin gene KRT82 SSRs (magenta highlight) are nearly conserved in the orthologues, except that they are longer in the macaque sequence (**Figure 12**). KRT85 SSR is shorter than orangutan repeat and longer than macaque repeat in intron VII.

Human KRT86 Intron I (ab-initio X and XII) tetranucleotide repeat (ACAG)n is longer in the human gene and the first mononucleotide repeat (T)n is partially (TA)n and partially (T)n in chimp and shorter in orangutan sequences. Each nucleotide of pentanucleotide repeat (AAACA) is conserved but the second mononucleotide repeat (T)n is longer in chimp and shorter in orangutan gene sequences (**Figure 12**).

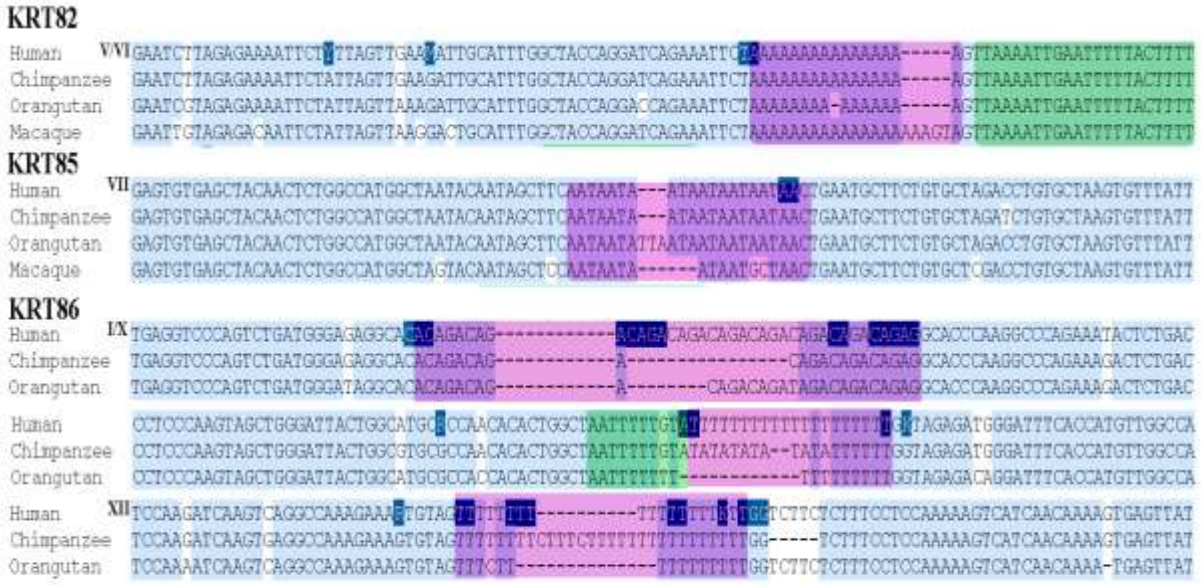

Fig.-12: SSRs (magenta highlight) in human hair Type II hair keratin genes and aligned sequences of chimp, orangutan and macaque orthologues. Some repeats are longer in orthologues compared to human repeats. Regions that are not SSRs but have repetitive nature (green highlight) are also seen in KRT82 and KRT86. (Aligned sequences from Ensembl[51,52])

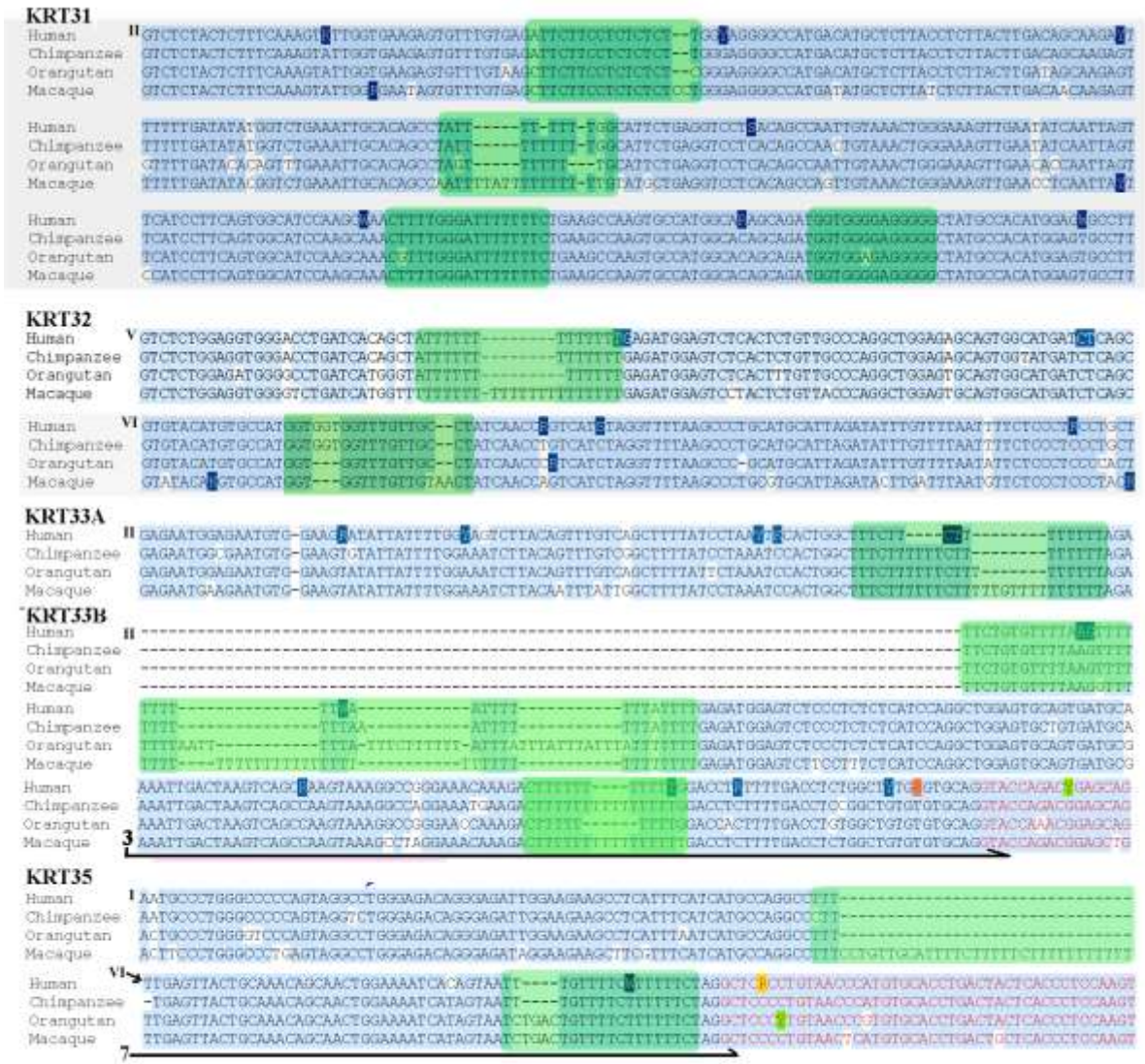

Fig.-13a: Repeat like regions (green highlight) in introns of Type I human hair keratin genes KRT31-KRT33B and KRT35. Repetitive regions are longer in intron II of KRT31, intron V of KRT32, intron II of KRT33A and intron I of KRT35 in macaque orthologues. Repeat like region is longer in intron II of KRT33B in orangutan and macaque orthologues. (Aligned sequences from Ensembl[51,52])

Fig.-13b: Repeat like regions (green highlight) in introns of Type I human hair keratin genes KRT36-KRT9. Repeat like regions are longer in intron I of KRT39 orthologue in chimp (yellow and green highlight) and in intron VI and exon 7 of KRT39 orthologue in macaque (green and yellow highlight). Aligned sequences from Ensembl[51,52])

Some sequence regions in Type I hair keratin genes show tendency towards repetitive pattern (green highlight) and can be marked for future investigation to monitor any alterations specifically in these regions (**Figure 13a** and **b**). Human KRT33A intron II shows sequence region that could have either (T)n or (TTC)n repeat in the orthologues. Thus, there is a reduction in repeat in human KRT33A intron II. A similar situation of possible reduction in two repeats in human KRT33B intron II compared with orthologues is also seen (**Figure 13a** and **b**). KRT39 sequence regions that show repetitive nature but are not SSR in intron I where (TTNTTN)n is seen only in the chimpanzee sequence (marked yellow); and (T)n repeat (not present in human) in area aligned region of exon 7 of human gene is seen only in the macaque sequence (marked yellow). Type II hair keratin genes also have repeat-like regions that are shown in **Figure 14**. Human KRT86 Intron I (ab-initio X and XII) has some regions that could be repetitive in nature (green highlight).

Fig.-14: Repeat like regions (green highlight) in introns of Type II human hair keratin genes KRT82, KRT85 and KRT86. Repeat like regions are longer in intron vii of human KRT85 and orthologue in chimp (green highlight). In intron I/X of KRT86 repeat like regions are either longer in human sequence or in orthologues or both. Exon 13 of KRT86 has longer repeat like region in orthologues (green highlight). Aligned sequences from Ensembl[51,52])

KERATIN GENES STRUCTURAL ORGANIZATION

Number of exons: Human hair keratin genes may have more than one transcript (**Table-1a** [2, 3, 6-9, 57-59] and **b** [2, 3, 8, 57-59]). Considering exon start-end co-ordinates and exon ranks in transcripts, total number of exons calculated in human hair keratin genes and the respective orthologues comparisons show similarities and exceptions. Number of exons are seven in Type I keratin genes, except eight in KRT35, KRT36 and KRT39 and nine exons in KRT40 genes. Human Type II hair keratin genes have nine exons, except eleven in KRT86 gene (**Table-2** [51, 52]). However, ab-initio (not empirical) exon numbers along with empirical exons are different, as these numbers are higher in some genes, evident from the marked up regions in sequences obtained from Ensembl 75 (data not shown), from the ones listed in **Table-2** [51, 52]. Human KRT39 has three transcript variants (**Table-1a** [2, 3, 6-9, 57-59]) and considering all variants, this gene has at least eight exons. Human KRT40 has ten exons.

Comparisons of human gene organizations with the chimp, orangutan and macaque orthologues show similar number of exons in Type I and Type II gene orthologues (if present). Platypus gene has only six exons compared with nine exons in all other KRT82. However, the ab-initio exons show a total of eleven exons in human KRT82 (data not shown). The differences in exon numbers could be due to transcript variants. Possibly because of at least four transcript variants, the human KRT86 ab-initio exon number is as high as 32 (data not shown) but the exons in graphical representation show 11 exons (data not shown).

Table-1a: Type I hair keratin genes names, characteristics, expression & pairing (information based on Ensembl[51,52] & Ref indicated in the table)

Hair Keratin	Alias	Strand	% GC	Transcript count	Expression	Expression Time	Pairing	Pairing In	Disease	Other	Reference
KRT31	KRTHA1, K31, Ha1	-1	49.81	1	Mid, upper cortex, medulla		K85	Lower cortex			2, 6, 8, 57, 58
KRT32	K32, KRTHA2, Ha2	-1	51.78	1	Cuticle, matrix, lower cortex	Much earlier than K39 & K40	K82 & K85	Mid cuticle			6
							K85	Lower cuticle			2,6-8,57,58
KRT33A	KRTHA3A, Ha3-I	-1	48.36	1	Mid & upper cortex, medulla	Simultaneous with K33B & K34			Polished mutation	Similar to K33b	2, 8, 57, 58,
KRT33B	KRTHA3B, Ha3-II	-1	46.12	1	Cortex & medulla	Simultaneous with K33A & K34		Mid & upper cortex		Similar to K33a	2, 8, 58
KRT34	K34, KRTHA4, Ha4	-1	49.68	1	Mid & or only upper cortex, medulla	Simultaneous with K33A & K33B	K86	Upper cuticle & cortex			2, 6-8, 57, 58
KRT35	K35, KRTHA5, Ha5	-1	53.66	2	Cuticle, precortex, lower, mid & upper cortex, matrix		K85	Matrix			2, 6-8, 57, 58
KRT36	KRTHA6, Ha6	-1	53.52	2	Mid & upper cortex, medulla	Where expression of K35 is finished	K81				2, 7, 9, 57-59
KRT37	KRTHA7, Ha7	-1	51.98	1	Mid & upper cortex		K81, K85 & K86				2, 3, 8, 57, 58
KRT38	KRTHA8, Ha8	-1	51.65	1	Mid & upper cortex, medulla						2, 8, 57, 58
KRT39	K39, Ka35	-1	42.38	3	Upper cuticle, cortex & medulla	After K32 expression	K82 & K86	Upper cuticle & cortex		Not >60% homology with k40	2, 7, 8, 58
KRT40	K40, Ha8	-1	41.71	3	Upper cuticle	After K32 expression	K82	Upper cuticle & cortex		Not >60% homology with k39	2, 6-8

Table-1b: Type II hair keratin genes names, characteristics, expression and pairing (informatin based on Ensembl[51,52] & Ref indicated in the table)

Hair Keratin	Alias	Strand	% GC	Transcript count	Expression	Expression Time	Pairing	Pairing In	Disease	Characteristics	Reference
KRT81	K8, Hb1	-1	58.16	1	Medulla, mid & upper cortex	Same time as K83			Monilethrix lymphoblastic leukemia MOLT-4	Similar to K83 & K86	2, 3, 57-59
KRT82	K82, Hb2	-1	52.58	1	Cuticle	Later than other genes in cuticle, precicely in lower	K32	Mid cuticle		Similar to K84 & K85	2, 3, 8, 57-59
							K39 & K40	Upper cuticle & cortex			
KRT83	K83, Hb3	-1	53.87	1	Medulla, mid & upper cortex	Along with K81 & K86			Monilethrix		2, 3, 57-59
KRT84	K84, Hb4	-1	50.99	1	Tongue					Similar to K82 & K86	2, 3, 59
KRT85	K85, Hb5	-1	54.51	3	Mid & upper cortex, matrix & precortex		K31	Lower cortex	Hair & nail ectodermal dysplasia		2, 3, 8, 57-59
							K32	cuticle, lower cuticle			
							K35	Matrix			
KRT86	K86, Hb6	1	49.75	4	Medulla, mid & upper cortex	After K81 or K83 in cortex	K34 & K39	Upper cuticle & cortex	Monilethrix		2, 3, 8, 57

Table-3: Comparison of exon lengths of human hair keratin genes in the orthologues

Length	Gene	Exon Numbers									
		1	2	3	4	5	6	7	8	9	10
Long	KRT31	Chimp									
	KRT32							Chimp			
	KRT33A	Chimp	Macaque					Chimp			
	KRT36		Chimp			Chimp					
			Macaque			Macaque					
			Orangutan			Orangutan					
	KRT37	Chimp						Chimp			
								Macaque			
	KRT38	Chimp						Chimp			
	KRT39	Chimp		Chimp			Chimp	Chimp	Chimp		
		Macaque		Macaque			Macaque	Macaque	Macaque		
				Orangutan			Orangutan				
	KRT40	Chimp	Chimp					Chimp			
		Macaque	Macaque				Macaque	Macaque			
		Orangutan	Orangutan								
	KRT81			Orangutan			Orangutan		Orangutan		
	KRT82	Chimp									
	KRT84				Orangutan						
	KRT85	Chimp	Chimp				Chimp				
		Macaque	Macaque				Macaque				
			Orangutan	Orangutan			Orangutan		Orangutan		
	KRT86	Chimp	Chimp		Chimp	Chimp	Chimp	Chimp	Macaque	Chimp	Macaque
		Macaque	Macaque				Macaque				
Short	KRT31	Orangutan					Orangutan				
	KRT32	Chimp									
		Macaque						Macaque			
		Orangutan						Orangutan			
	KRT33A	Orangutan						Orangutan			
	KRT33B	Orangutan						Orangutan			
	KRT34	Macaque						Macaque			
	KRT35							Chimp			
			Macaque					Macaque			
			Orangutan					Orangutan			
	KRT36	Chimp		Chimp			Chimp	Chimp			
		Macaque		Macaque			Macaque	Macaque			
		Orangutan		Orangutan			Orangutan	Orangutan	Orangutan		
	KRT37							Orangutan			
	KRT38							Macaque			
								Orangutan			
	KRT39		Chimp			Chimp					
			Macaque			Macaque					
		Orangutan	Orangutan			Orangutan		Orangutan	Orangutan		
	KRT40	Chimp		Chimp	Chimp	Chimp	Chimp	Chimp			
		Macaque		Macaque		Macaque	Macaque	Macaque			
		Orangutan		Orangutan		Orangutan	Orangutan	Orangutan			
	KRT81	Chimp									
		Orangutan	Orangutan		Orangutan	Orangutan		Orangutan		Orangutan	
	KRT82									Chimp	
		Macaque								Macaque	
		Orangutan								Orangutan	
	KRT83	Macaque								Macaque	
		Orangutan								Orangutan	
	KRT84	Chimp									
		Macaque									
		Orangutan	Orangutan	Orangutan		Orangutan	Orangutan				
	KRT85	Chimp		Chimp	Chimp	Chimp		Chimp			
		Macaque		Macaque	Macaque	Macaque		Macaque		Macaque	
		Orangutan			Orangutan	Orangutan		Orangutan		Orangutan	
	KRT86		Chimp	Chimp	Chimp		Chimp		Chimp		
		Macaque	Macaque	Macaque	Macaque	Macaque		Macaque		Macaque	Macaque

Table-2: Human hair keratin gene IDs, chromosome locations and number of exons (information based on Ensembl[51,52])

Human Gene ID	Chromosome	Associated Gene Name	Exons	Chimp ID	Chimp Chromosome	Exons	Orangutan ID	Orangutan Chromosome	Exons	Macaque ID	Macaque Chromosome	Exons	Platypus ID	Platypus Chromosome	Exons
ENSG00000094796	17	KRT31	7	ENSPTRG00000009161	17	7	ENSPPYG00000008425	17	7	ENSMMUG00000010756	16	7	ENSOANG00000004713	Contig13052	7
														Contig17464	8
ENSG00000108759	17	KRT32	7	ENSPTRG00000009164	17	7	ENSPPYG00000008420	17	7	ENSMMUG00000018644	16	7			
ENSG00000006059	17	KRT33A	7	ENSPTRG00000009158	17	7	ENSPPYG00000008427	17	7	ENSMMUG00000010754	16	7	ENSOANG00000004713	Contig13052	7
ENSG00000131738	17	KRT33B	7	ENSPTRG00000009159	17	7	ENSPPYG00000008426	17	7						
ENSG00000131737	17	KRT34	7							ENSMMUG00000010757	16	7		Contig17464	8
ENSG00000197079	17	KRT35	8	ENSPTRG00000023885	17	8	ENSPPYG00000008419	17	7	ENSMMUG00000018645	16	7			
ENSG00000126337	17	KRT36	8	ENSPTRG00000009165	17	8	ENSPPYG00000008418	17	8	ENSMMUG00000018646	16	8	ENSOANG00000007596	Contig4383	8
													ENSOANG00000007597	Contig4383	8
ENSG00000108417	17	KRT37	7	ENSPTRG00000009162	17	7	ENSPPYG00000008423	17	7	ENSMMUG00000018643	16	7	ENSOANG00000014895	Contig4383	7
ENSG00000171360	17	KRT38	7	ENSPTRG00000009163	17	7	ENSPPYG00000008421	17	7	ENSMMUG00000018643	16	7		Contig5328	8
ENSG00000196859	17	KRT39	8	ENSPTRG00000023017	17	7	ENSPPYG00000008446	17	7	ENSMMUG00000015110	16	7	ENSOANG00000002902	Contig607	8
ENSG00000204889	17	KRT40	9	ENSPTRG00000030928	17	7	ENSPPYG00000008445	17	7	ENSMMUG00000011422	16	7	ENSOANG00000002903	Contig607	7
ENSG00000205426	12	KRT81	9	ENSPTRG00000029732	12	9	ENSPPYG00000004541	12	11	ENSMMUG00000031925	11	11	ENSOANG00000012632	Contig15331	9
													ENSOANG00000012605	Contig1058	9
													ENSOANG00000012607	Contig1058	10
													ENSOANG00000012606		
ENSG00000161850	12	KRT82	9	ENSPTRG00000004976	12	9	ENSPPYG00000004545	12	9	ENSMMUG00000018218	11	9	ENSOANG00000012609	Contig1058	8
ENSG00000170523	12	KRT83	9	ENSPTRG00000004973	12	9	ENSPPYG00000004542	12	9	ENSMMUG00000031924	11	11	ENSOANG00000012632	Contig15331	9
													ENSOANG00000012605	Contig1058	9
													ENSOANG00000012607	Contig1058	10
													ENSOANG00000012606		
ENSG00000161849	12	KRT84	9	ENSPTRG00000004975	12	9	ENSPPYG00000014495	3_ran	6	ENSMMUG00000018217	11	9			
ENSG00000135443	12	KRT85	9	ENSPTRG00000004974	12	9	ENSPPYG00000004544	12	10	ENSMMUG00000018216	11	9	ENSOANG00000002471	Contig25428	7
ENSG00000170442	12	KRT86	11	ENSPTRG00000004972	12	9				ENSMMUG00000031925	11	11	ENSOANG00000012632	Contig15331	9
													ENSOANG00000012605	Contig1058	9
													ENSOANG00000012607	Contig1058	10
													ENSOANG00000012606		

Orthologue sequence comparison and exon length (s): Though the exon numbers in each keratin gene orthologue are similar, comparisons of exon lengths of human hair keratin genes with their orthologues indicate differences in lengths in many orthologues (**Tables-3** [51, 52]). KRT33B, KRT35, KRT83 do not have longer exons in any orthologue compared with human exons. Though there are sequence similarities indicated in the aligned sequence of human KRT34 gene with its orthologues, there is no structural similarity except with the macaque gene. Macaque exons 1 and 7 have shorter lengths compared with respective human exons (**Table-3** [51, 52]). Human KRT40 has ten exons but the first three exons and introns have no orthologue sequence in these regions (data not shown). No orthologue sequences of human KRT81 were available. Similarities of human hair keratin genes with platypus orthologues are not significant. Most introns in hair keratin genes begin with 5' GT and end with 3' AG nucleotide combination. The exceptions are introns I in KRT36 (3' ACC) and KRT37 (5' CT), but possibly due to ab-initio exon else other marking. Similarly, IX intron of KRT40 (5' GC), III intron KRT81 (5' GC), VI intron KRT85 (5' GG), KRT86 many introns including ab-initio do not have 5' GT and 3' AG pattern. KRT86 has CT nucleotide in 5' of ab-initio introns II, IV-IX, XI, XIV-XVIII, XX, XXI, XXX and XXXI. The other nucleotide types in 5' are TC (III), GG (XIII), TG (XXII) and GC (XXV). In the 3' end the nucleotide types are GC (I, XIX), AC (III-IX, XII, XIV-XVIII, XX, XXI, XXX), AT (X), GG XIII and TC (XXII and XXIX) (parenthesis have intron numbers).

DISCUSSION

SSRs are found in most organism genomes where repeats are present in coding and non-coding regions; and show differences in distribution, types, motifs, densities etc.[21, 22, 29, 35, 60-63]. The present study shows repeats in 70.59% hair keratin genes amongst which 63.64% are Type I genes and 83.33% genes are Type II genes. We find significant negative correlation between gene lengths and repeat density, unlike the positive correlation between repeats and genome size in mammals [22]. This however, does not imply contradiction with reports on genome size and repeat correlation, but indicates that repeat density may vary in different genes and may show characteristic deviations in the keratin genes from the overall genome pattern. It is possible that the human hair keratin genes have repeats for some functional reasons as studies have suggested importance of repeats in different genes [19, 23, 25, 30, 31, 38, 42-49, 64]. The present study shows few trinucleotide repeats; and dinucleotide repeats are not as abundant as tetra- and mono-nucleotide repeats. This is consistent with reported abundance of dinucleotide repeats but low frequency of trinucleotide repeats in human genome [22]. Except for mononucleotide repeat (T)n, most repeat motifs are not common amongst human hair keratin genes, which is like the inconsistency of repeats types observed between insect orders [65].

Abundance of mononucleotide runs, especially in introns of mammalian genomes, is considered a non-random process [19] and perhaps, mononucleotide repeats in human hair keratin genes are due to such process. Though mononucleotide repeats can be involved in gene regulation [44] and nucleosome positioning [45]; such functions are not known in human keratin genes having these repeats (KRT86, KRT82, KRT36, KRT39, KRT40, KRT31). Comparison of human hair keratin genes with Neanderthal genes may shed light on the hypothesis for fungal genomes which suggests presence of mononucleotide repeats in small genes as remains of genes that contained repeats or platform for formation of new gene [35]. Dinucleotide repeats like AT(n) increase DNA flexibility as well as association with basic proteins, and thus may help in gene regulation and chromatin dynamics [66]. Such functions of repeats in hair keratin genes need investigation. Though polymorphic sites in KRT32 and some hair keratin genes are not within or near SSRs, this indicates a possibility of mutations in these genes as the propensity of microsatellites to mutate is higher compared to other sequence regions especially in coding sequences [22, 67].

This suggestion gets support from the study on KRTHAP1 (human keratin pseudogene), where base substation C/T in exon 4 resulted in TGA premature codon rendering the gene inactive, though the counterpart in chimp is an active gene [68]. Though Ha2 (KRT32) exons 6 and the intron 6 boundary have polymorphic sites [69], no SSR is present in this exon. However, the intronic repeat TTTAT (n) found in the present analysis is near region that is known for intronic variations as per Ensembl version 75.

Repeats in exons/introns: Though similar mechanisms operate in repeat expansion in coding and non-coding sequences [70], the intronic repeats have lower selection pressure compared with repeats in exons [23]. Few repeats in coding regions are perhaps to avoid possibility of frame-shift or other mutations that may affect gene expression. Repeats in exons are possibly those that have functional roles [22, 23, 32, 71-74]. Repeat expansion in coding regions can alter protein function which can also result in new function [36]. It is perhaps due to these reasons that presence of SSRs in exons compared with introns is low in the present study, which corroborates studies on different genes and genomes [22, 28, 62, 72, 75]. Some trinucleotide repeats may add new coding regions and affect size or add new functions to proteins [76]. It is possibly due to this reason that triplet and hexa repeats are present only in introns of the genes in the present study. Paradoxically, this is unlike observations of abundance of tri- and hexa-nucleotide repeats in whole human genome [77]. Thus, distributions of tri- and hexa-nucleotide repeats appear to be gene specific and differ from distribution pattern seen in

whole genome. Triplet repeat expansions can also lead to non-ATG (RAN) translation i.e. can initiate translation without ATG codon [41]. Therefore, even if repeats are present in introns, there is a possibility of RAN translation. This may be the reason for presence of few repeats in human hair keratin genes. Further, repeats are more frequent in newly formed introns compared with established introns and may play role in evolution [78]. In the present study, many repeats are present in or near regions mapped for intronic variations. For example, mononucleotide repeat T(n) in intron of KRT31 and A(n) repeat in KRT82 lie in the region where intronic variants have been found. Intronic variations can result in allele variations [79] and quantitative trait variations [80]. Possibly, SSRs may be one of the contributory agents for polymorphism, which may include variations (if any) in expression of the hair keratin genes. Tough repeats may have functional roles, SSRs instability can lead to disorders or interruption in gene functions or disease susceptibility [23, 32, 34, 40]. Even intronic repeat mutation can be one of the causes for disorders [20, 36, 81, 82] or affect promoter activation [40, 83]. It remains to be seen whether repeats found in the present study are susceptible for mutations that may alter functions of genes. Since repeats are found in the introns of KRT81, KRT83 and KRT86 genes, SSR mutations cannot be associated with point mutations that cause monilethrix and KRT85 mutations that affects hair matrix as these mutations are in exons involving change in amino acid [2, 9, 84]. It has been reported that stabilized genes have lower mutation frequencies especially in introns compared with paralogues or orthologues [78]. Possibly, Type I keratin genes KRT33A, KRT35, KRT37 and KRT38 and Type II KRT84 gene have also become stable and thus have no SSRs. This suggestion gets support from reported higher sequence stability of KRT33A compared with KRT31 and KRT33B where the latter two genes share closer neighbourhood phylogenetic relationship [8, 85]. However, this argument does not stand true in case of KRT35, which is a closer neighbour of KRT32 protein [8, 85], but this gene does contain repeats. It is possible that the differences in these genes are due to differences in non-keratin protein immediate interacting partners, or different selection pressures acting on these genes.

Repeat CG richness: The present study does not show significant correlation between CG richness of gene sequence and repeat CG richness. Further, the repeats do not have greater than 75% CG richness, which agrees with other studies where the GC rich repeats are not abundant in genomes, particularly in coding sequences [23, 48, 74, 86]. Few C/G rich repeats in human keratin genes could be due to the propensity for mutations [28, 74, 87-89] even though some mammalian genes have higher ratio of C/G repeats of certain lengths [74]. Human KRT31 exon has a tetranucleotide repeat CTCC(n) which lies in the region that is known for variation as per the Ensembl version 75. This indicates a possibility of this repeat affecting regulatory functions in KRT31, as is found in case of other genes [23].

Repeat length: Longer repeats are polymorphic [27], have higher chances of mutations interrupting the length [65], tend to have higher slippage rates [28] and are deleted in dividing cells [70]. Since the hair keratin genes are also expressed in actively dividing cells (the hair bulb), it is possibly due to this reason and propensity for higher slippage rate, that most repeats are not long in the hair keratin genes. Shorter repeats in human hair keratin genes compared with orthologues could be interpreted as a decay of repeats.

However, unlike studies on insect orders [65], this cannot be assumed to reflect the rate of evolution as some repeats are longer in the human hair keratin genes (e.g. KRT86 tetranucleotide repeat) compared with orthologues, which agrees with reported longer repeats in humans compared with chimpanzee [27]. Repeat length difference could indicate differences in mutation rates in different species and mutational bias toward either repeat expansion or contraction. Thus, different mechanisms and pressures must be acting that result in differences in repeats and their conservation in different organisms [65], and on different genes even though they are involved largely in the same function in the same species like the present study.

Conservation of Repeats: Conservation of SSRs in non-coding regions may indicate some functional importance e.g., in gene regulation [25], selection or stability of genes [29] or facilitators of evolutionary adaptations by providing adjustability [25]. In light of these functions; lowest percentage of repeat conservation despite presence of many SSRs in KRT86 gene in the present study indicates possibility of further evolutionary alterations due to mutations in SSRs. KRT81, KRT83 and KRT86 differ in their repeat types and densities notwithstanding their conservation in respective orthologues. The differences in repeat conservation is despite close phylogenetic relationship of keratin Type II proteins KRT86, KRT81 and KRT83 proteins [8, 85]. Further, disparity in repeats found in these genes is intriguing despite their expression in the same regions of hair and possibly dimerisation with similar Type I keratins [5, 6] (**Table-1a** [2, 3, 6-9, 57-59] and **b** [2, 3, 8, 57-59]). KRT39 and KRT40 show differences in repeat types and their conservation, where KRT39 has less diversity of repeat motifs compared with KRT40.

Despite same evolutionary branching, these two genes do not have >60% sequence similarity with each other and other members of Type I keratin; their expression patterns are different and form dimers with different Type II genes [5, 6, 8] (**Table-1a** [2, 3, 6-9, 57-59] and **b** [2, 3, 8, 57-59]). It would be interesting to investigate whether the repeat differences in these genes have any contribution in either their evolutionary divergence, regulation of expression and or association with respective Type II keratins. Comparisons of SSRs in human genes with orthologues indicate that some SSRs may have same location but the motif may change in different species [25]. This is similar to findings in the present study where KRT86 human gene has (T)n mononucleotide repeat but the orthologue in chimp has partial (TA)n and partial (T)n repeat in the same location as per the aligned sequence (**Figure 12**). Repeats in regulatory regions like promoter sequences and their conservation also indicates importance of repeats in gene regulation and possibly phenotypic variations in humans [45]. Repeat (CTCC)n in UTR of KRT31 gene in the present study and its conservation when compared with three primates indicates a possibility of involvement of this repeat in regulatory functions. On the other hand, conservation of repeats across different species may or may not have functional importance. Instead, SSRs could be conserved due to their neutral mutations. Human specific repeats as found in the present study indicate species and gene specific association of SSRs.

STRUCTURAL ORGANIZATION

Study of structural organization of exons/introns in cytoplasmic IF protein genes shows homologous position of many introns; indicating common ancestry of lamins and IF proteins [13, 90]. Variations and conservations of introns in lamin genes indicate linage specificity [90]. The present study shows similarity in structural organization of human hair keratin genes of the same Type, but Type I and Type II genes are not similar. This is in agreement with earlier studies on keratin genes that show similarity in sequences, exon/intron structure including Type I hair keratin genes [8] and conserved elements in promoter region [10]. Analyses of introns in most eukaryotic genes also confirm conservation of positions of introns [13, 91-93]. The present study shows partial concurrence with the study that suggests higher intronic burden in evolutionarily conserved genes [94]. Nevertheless, KRT86 has highest number of exons (ab-initio) and thus highest number of introns compared with other hair keratin genes and does not show conservation of many regions with orthologues. This indicates a possibility of further variations in this gene in future. In concurrence with studies on eukaryotic genes [95], the first exons/introns in most hair keratin genes show relatively higher conservation, although there are length differences. For example, the first exons of many orthologues in chimp are longer. On the other hand, orangutan orthologue shows longer first exon only in Type I keratin KRT40 but shorter first exon in many Type II keratin genes.

In case of KRT40, first three exons of all three primate orthologues do not show similar lengths. Though no orthologue sequence alignments of KRT81 were available, the exon start-end co-ordinates of orthologues indicate that only macaque gene first exon is similar with human exon length. Further, besides the first exon/intron, there are differences in lengths of other exons/introns in orthologues of human hair keratin genes. However, these differences could be also due to annotation differences [96]. SSRs may have some role in length differences at least in some exons/introns as some repeats in these regions show differences in length compared with human repeats. However, the differences in exon/intron lengths and the role if any of SSRs in these dissimilarities need further investigation. Further, conservation of exons/introns and repeats in them in the present study suggest functional importance of at least some introns [91-93] and therefore of the SSRs within them. Many human repeats correspond to same gene regions as those of the orthologues, whereas some repeats are longer in human genes, others are shorter. This indicates the dynamic nature of SSRs that are undergoing alterations in human genes compared with the orthologue genes. Because conserved SSRs are possible indications of selection or stability of genes [29], conservation of many repeats and no significant alteration in gene structure in the present study indicates common ancestry of these repeats.

Repeats have also played role in exon-intron duplication especially in segmental duplication [31, 97]. It is possible that repeats in hair keratin genes could have contributed in evolution and divergence of these genes from non-human primate genes and other mammals besides playing a role in structural organization. Repeats may also play a role in primate specific and particularly human specific segmental duplication, enabling new expression that may be tissue specific or facilitate new functions to the protein coding genes where they are present. This may be achieved by elongation of repeats leading to longer amino acid repeats thus, modifying the protein [97]. Most introns in hair keratin genes begin with 5’ GT and end with 3’ AG nucleotide combination which is similar to observations of study on keratin Type II gene [13].

However, there are some exceptions that could be due to ab-initio exon markings. Significance of this finding remains elusive. SSRs in hair keratin genes need investigation to seek variations with reference to human populations living in diverse climatic conditions. Though Neanderthal hair keratins have been compared with human genes [98]; a comparison based on SSRs has not been done. Such investigations may help in further strengthening the hypothesis that SSR variability may help in adaptations under changing environmental conditions with reference to expression of hair keratin genes. Disparities in repeats in similar keratin genes yet conservation of repeats in many genes need further investigation from point of view of their evolutionary divergence. This is pertinent when studies show that human hair keratin genes are amongst genes undergoing positive selection and adaptive evolution.

These genes have higher allele frequency differences, which may affect long term adaptive changes in response to changing environment conditions [99]. Neanderthal keratin genes (including hair keratin) could have helped humans adapt to non-African environment[98]. Analysis of SSRs in Neanderthal genes would help in understanding the roles of SSRs (if any) in alterations in genes between Neanderthal and modern humans. Alterations in gene structure or gene polymorphism could also have association with repeat conservation as they could be involved in intron removal. Removal of intron would lead to alteration in gene structure and function of the gene product [100].

CONCLUSION

Human hair keratin genes are amongst evolving genes under positive selection pressure [99]; the changes in

these genes are possibly less dependent on hypermutable SSRs. This inference is based on the fact that many human keratin genes analysed in the present study do not have repeats or have few repeats. Moreover, except for one, all repeats are present in introns.

Any alterations in repeats in introns are less likely to affect gene functions compared with repeats in exons. Conversely, many of these repeats are present in regions that are known for intronic variations. Thus, though SSRs, which can cause alterations in genes and possibly lead to evolutionary changes, are not abundant, they may play an important role. Only two hair keratin genes have >50% CG rich repeat motifs. Thus, possibilities of mutations especially due to strand slippage that are higher in C/G rich sequences are less likely in these genes. Platypus repeats show least conservation and chimpanzee repeats show highest conservation, as expected. Conservation of many human repeats in the present study indicates possible functional importance of these SSRs. Some repeats are human specific, as they do not show conservation when compared with orthologues in the four mammals. These repeats may have possibly originated after the divergence of human genes from ancestral mammalian genes. The structural organization of human hair keratin genes is largely similar but Type I and Type II are not similar. Though there is a general similarity and conservation of gene organization, some exons are either shorter or longer in the orthologues compared to human exons/introns. SSRs may have some role in these length differences as some repeats in hair keratin genes show differences in length compared with human repeats. The differences in exon/intron lengths and the role if any of SSRs in these dissimilarities need further investigation. KRT86, which has many repeats, also shows little similarity in structural organization. It is possible that at least in this gene, repeats are involved in alterations in the gene architecture, which is the probable reason for deviation from the structural organization of Type II hair keratin genes within human genome and with the orthologues. Platypus orthologue sequences do not show significant similarity with human genes. It is possible that the human specific repeats in hair keratin genes could have contributed in evolution and divergence of these genes from the non-human primate genes and other mammals, besides playing a role in structural organization of these genes. Analysis of SSRs in Neanderthal genes would help in understanding the roles of SSRs (if any) in alterations in genes between Neanderthal and modern humans.

ACKNOWLEDGMENT

We are grateful to Dr Miroslav Blumenberg (Associate Professor, Departments of Dermatology and Biochemistry and Molecular Pharmacology, New York University, USA) for his suggestions for improving the manuscript and to Mr Abhay Vinayak Pendse for help with computational work.

REFERENCES

1. G.E. Rogers. *Int. J. Dev. Biol.* 2004, **48**, 163-70.
2. R. Moll, M. Divo and L. Langbein. *Histochem. Cell Biol.* 2008, **129**, 705-33.
3. H.H. Bragulla and D.G. Homberger. *J. Anat.* 2009, **214**, 516-59.
4. J. Schweizer, P.E. Bowden, P.A. Coulombe, L. Langbein, E.B. Lane, T.M. Magin, L. Maltais, M.B. Omary, D.A.D. Parry, M.A. Rogers and M.W. Wright. *J. Cell Biol.* 2006, **174**, 169-74.
5. L. Langbein, H. Yoshida, S. Praetzel-Wunder, D.A. Parry and J. Schweizer. *J. Invest. Dermatol.* 2010, **130**, 55-73.

6. L. Langbein, L. Eckhart, M.A. Rogers, S. Praetzel-Wunder and J. Schweizer. *J. Biol. Chem.* 2010, **285**, 36909-21.
7. L. Langbein, M.A. Rogers, H. Winter, S. Praetzel, U. Beckhaus, H.R. Rackwitz and J. Schweizer. *J. Biol. Chem.* 1999, **274**, 19874-84.
8. L. Langbein, M.A. Rogers, S. Praetzel-Wunder, D. B?ckler, P. Schirmacher and J. Schweizer. *J. Invest. Dermatol.* 2007, **127**, 1532-1535.
9. Y. Honda, K. Koike, Y. Kubo, S. Masuko, Y. Arakawa and S. Ando. *Cell Struct. Funct.* 2014, **39**, 31-43.
10. B. Powell, L. Crocker and G. Rogers. *Development*. 1992, **114**, 417-33.
11. M.A. Rogers, H. Winter, L. Langbein, T. Krieg and J. Schweizer. *J. Invest. Dermatol.* 1996, **107**, 633-8.
12. M.A. Rogers, H. Winter, C. Wolf, J. Schweizer and L. Langbein. *J. Invest. Dermatol.* 2000, **114**, 464-472.
13. E.M. Klinge, Y.R. Sylvestre, I.M. Freedberg and M. Blumenberg. *J. Mol. Evol.* 1987, **24**, 319-329.
14. J.C. Chamcheu, I.A. Siddiqui, D.N. Syed, V.M. Adhami, M. Liovic and H. Mukhtar. *Arch. Biochem. Biophys.* 2011, **508**, 123-137.
15. W.H.I. McLean and C.B.T. Moore. *Hum. Mol. Genet.* 2011, **20**, R189-R197.
16. J. Wu, Y. Lin, W. Xu, Z. Li and W. Fan. *J. Biomed. Res.* 2012, **25**, 49-55.
17. G. Avhad and P. Ghuge. *Int. J. Trichology*. 2013, **5**, 224.
18. Y. Shimomura, M. Wajid, M. Kurban, N. Sato and A.M. Christiano. *J. Invest. Dermatol.* 2010, **130**, 892-895.
19. G. Levinson and G.A. Gutman. *Mol. Biol. Evol.* 1987, **4**, 203-221.
20. V. Campuzano, L. Montermini, M.D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S.I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P.I. Patel, S. Di Donato, J.L. Mandel, S. Cocozza, M. Koenig and M. Pandolfo. *Science.* 1996, **271**, 1423-1427.
21. G.K. Chambers and E.S. MacAvoy. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* 2000, **126**, 455-476.
22. H. Ellegren. *Nat. Rev.* 2004, **5**, 435-445.
23. Y.C. Li, A.B. Korol, T. Fahima and E. Nevo. *Mol. Biol. Evol.* 2004, **21**, 991-1007.
24. G.F. Richard, A. Kerrest and B. Dujon. *Microbiol. Mol. Biol. Rev.* 2008, **72**, 686-727.
25. D.G. King. *In Advances in Experimental Medicine and Biology*. Springer New York, 2012, 10-25.
26. M. Litt and J.A. Luty. *Am. J. Hum. Genet.* 1989, **44**, 397-401.

27. W. Amos, *BioEssays*. 2010, **32**, 82-90.

28. G. Ananda, E. Walsh, K.D. Jacob, M. Krasilnikova, K.A. Eckert, F. Chiaromonte and K.D. Makova. *Genome Biol. Evol.* 2013, **5**, 606-20.

29. M.L. Szpara, D. Gatherer, A. Ochoa, B. Greenbaum, A. Dolan, R.J. Bowden, L.W. Enquist, M. Legendre and A.J. Davison. *J. Virol.* 2014, **88**, 1209-27.

30. Y. Kashi and D.G. King. *Trends Genet.* 2006, **22**, 253-259.

31. H. Keren, G. Lev-Maor and G. Ast. *Nat. Rev. Genet.* 2010, **11**, 345-355.

32. J.M. Hancock and M.F. Santibáñez-Koref. *EMBO J.* 1998, **17**, 5521-4.

33. C. Schlötterer. *Chromosoma*. 2000, **109**, 365-371.

34. M.I. McCarthy. *Hum. Mol. Genet.* 2004, **13**, 33R-41.

35. M.W. van Passel and L.H. de Graaff. BMC Genomics. 2008, **9**, 596.

36. J.R. Brouwer, R. Willemsen and B.A. Oostra. *BioEssays*. 2009, **31**, 71-83.

37. D.K. Jeppesen, V.A. Bohr and T. Stevnsner. *Prog. Neurobiol.* 2011, **94**, 166-200.

38. A.J. Hannan. *In Advances in experimental medicine and biology*. Springer New York. 2012, **769**, 1-9.

39. H. Budworth and C.T. McMurray. *Methods Mol. Biol.* 2013, **1010**, 3-17.

40. J.S. Yokoyama, D.W. Sirkis and B.L. Miller. *Am. J. Neurodegener. Dis.* 2014, **3**, 1-18.

41. J.D. Cleary and L.P.W. Ranum. *Hum. Mol. Genet.* 2013, **22**, R45-51.

42. S.J. Tang. *Genes (Basel).* 2011, **2**, 502-515.

43. S.J. Tang. *Genes (Basel).* 2012, **3**, 167-175.

44. C. Aporntewan, P. Pin-on, N. Chaiyaratana, M. Pongpanich, V. Boonyaratanakornkit and A. Mutirangura. *Nucleic Acids Res.* 2013, **41**, 8872-8885.

45. S.M. Sawaya, A.T. Bagshaw, E. Buschiazzo and N.J. Gemmell. *Adv. Exp. Med. Biol.* 2012, **769**, 41-54.

46. R.J. Haasl, R.C. Johnson and B.A. Payseur. *Genome Biol. Evol.* 2014, **6**, 1843-61.

47. M.A. Varela and W. Amos. *Genomics*. 2010, **95**, 151-159.

48. C. Mayer, F. Leese and R. Tollrian. *BMC Genomics*. 2010, **11**, 277.

49. M.A. Huntley and A.G. Clark. *Mol. Biol. Evol.* 2007, **24**, 2598-2609.

50. J.W. Fondon and H.R. Garner. *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**, 18058-18063.

51. R.J. Kinsella, A. Kähäri, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey and P. Flicek. *Database (Oxford).* **2011**, bar030.

52. P. Flicek, M.R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C.G. Girón, L. Gordon, T. Hourlier, S. Hunt, N.

Johnson, T. Juettemann, A.K. Kähäri, S. Keenan, E. Kulesha, F.J. Martin, T. Maurel, W.M. McLaren, D.N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H.S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S.J. Trevanion, A. Vullo, S.P. Wilder, M. Wilson, A. Zadissa, B.L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T.J.P. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D.R. Zerbino and S.M.J. Searle. *Nucleic Acids Res.* 2014, **42**, D749-D755.

53. R. Kofler, C. Schlötterer and T. Lelley. *Bioinformatics*. 2007, **23**, 1683-1685.
54. G.M. Cooper, E.A. Stone, G. Asimenos. NISC Comparative Sequencing Program. E.D. Green, S. Batzoglou and A. Sidow. *Genome Res.* 2005, **15**, 901-913.
55. B. Paten, J. Herrero, K. Beal, S. Fitzgerald and E. Birney. *Genome Res.* 2008, **18**, 1814-1828.
56. P. Pavlidis and W.S. Noble. *Bioinformatics*. 2003, **19**, 295-6.
57. L. Langbein, M.A. Rogers, S. Praetzel, H. Winter and J. Schweizer. *J. Invest. Dermatol.* 2003, **120**, 512-522.
58. L. Langbein, M.A. Rogers, S. Praetzel-Wunder, B. Helmke, P. Schirmacher and J. Schweizer. *J. Invest. Dermatol.* 2006, **126**, 2377-2386.
59. L. Langbein, M.A. Rogers, H. Winter, S. Praetzel and J. Schweizer. *J. Biol. Chem.* 2001, **276**, 35123-35132.
60. G. Tóth, Z. Gáspári and J. Jurka. *Genome Res.* 2000, **10**, 967-81.
61. M.V Katti, P.K. Ranjekar and V.S. Gupta. *Mol. Biol. Evol.* 2001, **18**, 1161-7.
62. F. Cruz, M. Pérez and P. Presa. *Gene*. 2005, **346**, 241-247.
63. S. Trivedi. *Genet. Mol. Res.* 2006, **5**, 741-72.
64. C. Lenz, W. Haerty and G.B. Golding. *Genome Biol. Evol.* 2014, **6**, 655-665.
65. E. Stolle, J.H. Kidner and R.F.A. Moritz. *Genome Biol. Evol.* 2013, **5**, 151-162.
66. T.M. Okonogi, S.C. Alley, A.W. Reese, P.B. Hopkins and B.H. Robinson. *Biophys. J.* 2002, **83**, 3446-3459.
67. Y.C. Li, A.B. Korol, T. Fahima, A. Beiles and E. Nevo. *Mol. Ecol.* 2002, **11**, 2453-2465.
68. H. Winter, L. Langbein, M. Krawczak, D.N. Cooper, L.F. Jave-Suarez, M.A. Rogers, S. Praetzel, P.J. Heidt and J. Schweizer. *Hum. Genet.* 2001, **108**, 37-42.
69. H. Winter, M.A. Rogers, B. Mahr, M. Cremer, T. Krieg, H.A. Navsaria, I.M. Leigh and J. Schweizer. *J. Invest. Dermatol.* 1996, **106**, 544-8.
70. C.T. McMurray. *Nat. Rev. Genet.* 2010, **11**, 786-799.
71. C. Schlötterer. *Curr. Biol.* 1998, **8**, R132-R134.
72. N.V. Dokholyan, S.V. Buldyrev, S. Havlin and H.E. Stanley. *J. Theor. Biol.* 2000, **202**, 273-282.

73. D. Metzgar, J. Bytof and C. Wills. *Genome Res.* 2000, **10**, 72-80.

74. T. Gu, S. Tan, X. Gou, H. Araki and D. Tian. *Genetics.* 2010, **186**, 1077-84.

75. S. Subramanian, R.K. Mishra and L. Singh. *Genome Biol.* 2003, **4**, R13.

76. H. Green and N. Wang. *Proc. Natl. Acad. Sci. U. S. A.* 1994, **91**, 4298-302.

77. S. Subramanian, V.M. Madgula, R. George, R.K. Mishra, M.W. Pandit, C.S. Kumar and L. Singh. *Bioinformatics*. 2003, **19**, 549-52.

78. W. Li, R. Kuzoff, C.K. Wong, A. Tucker and M. Lynch. *Genome Biol. Evol.* 2014, **6**, 2218-2234.

79. L.F. Stam and C.C. Laurie. *Genetics*. 1996, **144**, 1559-64.

80. M. Farrall. *Hum. Mol. Genet.* 2004, **13**, 1R-7.

81. N. Cherng, A.A. Shishkin, L.I. Schlager, R.H. Tuck, L. Sloan, R. Matera, P.S. Sarkar, T. Ashizawa, C.H. Freudenreich and S.M. Mirkin. *Proc. Natl. Acad. Sci. U. S. A.* 2011, **108**, 2843284-8.

82. K.N. McFarland, J. Liu, I. Landrian, R. Gao, P.S. Sarkar, S. Raskin, M. Moscovich, E.M. Gatto, H.A.G. Teive, A. Ochoa, A. Rasmussen and T. Ashizawa. *Eur. J. Hum. Genet.* 2013, **21**, 1272-1276.

83. X. Guan, Z. Liu, L. Wang, L.E. Wang, E.M. Sturgis and Q. Wei. *Carcinogenesis*. 2013, **34**, 812-817.

84. M.A.M. van Steensel, P.M. Steijlen, R.S. Bladergroen, M. Vermeer and M. van Geel. *J. Med. Genet.* 2005, **42**, e19-e19.

85. M. Hesse, T.M. Magin and K. Weber. *J. Cell Sci.* 2001, **114**, 2569-75.

86. S. Trivedi. *Indian J. Biotechnol.* 2004, **3**, 331-347.

87. H. Moore, P.W. Greenwell, C.P. Liu, N. Arnheim and T.D. Petes. *Proc. Natl. Acad. Sci.* 1999, **96**, 1504-1509.

88. M.J. Ireland, S.S. Reinke and D.M. Livingston. *Genetics*. 2000, **155**, 1657-65.

89. N.L.S. Chan, J. Guo, T. Zhang, G. Mao, C. Hou, F. Yuan, J. Huang, Y. Zhang, J. Wu, L. Gu and G.M. Li. *J. Biol. Chem.* 2013, **288**, 15015-15022.

90. A. Peter and R. Stick. *Nucleus*. **3**, 44-59.

91. A. Rogozin, A.V Sverdlov, V.N. Babenko and E.V. Koonin. *Brief. Bioinform.* 2005, **6**, 118-134.

92. L. Carmel, I.B. Rogozin, Y.I. Wolf and E.V Koonin. *BMC Evol. Biol.* 2007, 7, 192.

93. L. Carmel, I.B. Rogozin, Y.I. Wolf and E.V Koonin. *Genome Res.* 2007, **17**, 1045-50.

94. O. Gorlova, A. Fedorov, C. Logothetis, C. Amos and I. Gorlov. *BMC Evol. Biol.* 2014, **14**, 50.

95. S.G. Park, S. Hannenhalli and S.S. Choi. *BMC Genomics*. 2014, **15**, 526.

96. S.W. Roy and D. Penny. *Genome Res.* 2006, **16**, 1270-1275.

97. A. De Grassi and F.D. Ciccarelli. *Genome Biol.* 2009, **10**, R137.

98. N. Patterson, D. Reich, S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson and D. Reich. *Nature*. 2014, **507**, 354-7.

99. R.D. George, G. Mcvicker, R. Diederich, S.B. Ng, A.P. Mackenzie, W.J. Swanson, J. Shendure and J.H. Thomas. DOI:10.1101/gr.121327.111.

100. H. Naora and J.N. Deacon. *Proc. Natl. Acad. Sci. U. S. A.* 1982, **79**, 6196-6200.

***Corresponding author: Seema Trivedi;** Department of Zoology, JN Vyas University, Jodhpur (Rajasthan), India.

On line publication Date: 08.08.2017