Ohadi *et al. Biology Direct* (2024) 19:70
https://doi.org/10.1186/s13062-024-00508-8

Biology Direct

RESEARCH

Open Access

Novel crossover and recombination hotspots massively spread across primate genomes

Mina Ohadi[1*], Masoud Arabfard[2*], Safoura Khamse[1], Samira Alizadeh[1], Sara Vafadar[1], Hadi Bayat[1,3], Nahid Tajeddin[1,4], Ali M. A. Maddi[5], Ahmad Delbari[1] and Hamid R. Khorram Khorshid[6]

Abstract

Background The recombination landscape and subsequent natural selection have vast consequences forevolution and speciation. However, most of the crossover and recombination hotspots are yet to be discovered. We previously reported the relevance of C and G trinucleotide two-repeat units (CG-TTUs) in crossovers and recombination.

Methods On a genome-wide scale, here we mapped all combinations of A and T trinucleotide two-repeat units (AT-TTUs) in human, consisting of AATAAT, ATAATA, ATTATT, TTATTA, TATTAT, and TAATAA. We also compared a number of the colonies formed by the AT-TTUs (distance between consecutive AT-TTUs < 500 bp) in several other primates and mouse.

Results We found that the majority of the AT-TTUs (> 96%) resided in approximately 1.4 million colonies, spread throughout the human genome. In comparison to the CG-TTU colonies, the AT-TTU colonies were significantly more abundant and larger in size. Pure units and overlapping units of the pure units were readily detectable in the same colonies, signifying that the units were the sites of unequal crossover. We discovered dynamic sharedness of several of the colonies across the primate species studied, which mainly reached maximum complexity and size in human.

Conclusions We report novel crossover and recombination hotspots of the finest molecular resolution, massively spread and shared across the genomes of human and several other primates. With respect to crossovers and recombination, these genomes are far more dynamic than previously envisioned.

Keywords Human, AT trinucleotide, Two-repeat, Unequal crossover, Recombination hotspot: primate, Shared

*Correspondence:
Mina Ohadi
mi.ohadi@uswr.ac.ir
Masoud Arabfard
arabfard@gmail.com
[1] Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
[2] Chemical Injuries Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
[3] Biochemical Neuroendocrinology, Montreal Clinical and Research Institute (IRCM, affiliated to the McGill University, Montreal, QC H2W 1R7, Canada
[4] Department of Biology, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[5] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[6] Personalized Medicine and Genometabolomics Research Center, Hope Generation Foundation, Tehran, Iran

Background

Crossover and recombination, alongside mutation, generate the raw material of evolution and speciation [1, 2]. Recombination hotspots are regions in a genome that exhibit elevated rates of recombination relative to a neutral expectation. Studies on recombination hotspots are mainly founded on mapping crossover events through pedigree analysis and linkage disequilibrium [3, 4]. Identification of these hotspots paved the way for the discovery of PRDM9, a trimethyl transferase, which is associated with hotspot activity in both humans and mouse [5–7]. Using HapMap data, Myers et al. identified a 13-bp "core" motif "CCTCCCTNNCCAC" for PRDM9 binding, which is strongly correlated with hotspot activity when it occurs in both repeat and non-repeat DNA. A close match to this motif was reported to occur in about 40% of the crossover hotspots known to date [8]. Degenerate versions of the motif, of variable binding activity for PRDM9, have since been identified in the human genome on centiMorgan (cM) scales [9, 10]. The 13-mer motif is the most characterized hotspot locus in human to date. However, the level of expression of PRDM9 should control for only a fraction of the targets that are hotspots and the overall temperature of the genome [11].

Other indirect approaches, such as phylogenetic and integrated genetic versus physical map analyses, led to the idea that the local rates of recombination are positively correlated with GC content in the human genome [12–15] and a few other mammals [16]. Lined with the above, there are reports that meiotic recombination favors GC- over AT-rich alleles, and facilitates local GC-content [17, 18]. When a meiotic recombination hotspot from a GC-rich isochore was inserted into an AT-rich isochore domain, the site adopted the lower recombination activity, characteristic of its new environment [19]. It is reported that programmed in vitro double strand break formation and loading of axial structure proteins are much more prominent in GC-rich isochores [9, 10, 12].

We previously reported that C and G trinucleotide two-repeat units (CG-TTUs) form colonies of exceeding significance across the human genome, based on Poisson distribution [20, 21]. Several of the large and medium size colonies that were further analyzed in other species, unveiled crossover and recombination hotspots, shared across primates, and in some instances, even in mouse.

Here, we investigated A and T trinucleotide two-repeat units (AT-TTUs) with a similar algorithm, and discovered that the colonies formed by AT-TTUs were significantly more abundant and larger than the colonies formed by CG-TTUs. These novel crossover sites vastly spread across primate genomes, and mainly reach maximum complexity and size in human.

Materials and methods

Whole-genome extraction of AT-TTUs in human

A Java software package was created (available at: https://github.com/arabfard/Java_STR_Finder) to facilitate the extraction of AT-TTUs, including AATAAT, ATAATA, ATTATT, TTATTA, TATTAT, and TAATAA, along with their corresponding locations (Fig. 1). To that end, we utilized the latest version of the human genome assembly (GRCh38. p14), obtained from the UCSC genome browser (accessible at https://hgdownload.soe.ucsc.edu).

To ensure the accuracy and reliability of the data obtained from the algorithm, a validation process was conducted. This involved random manual examination of

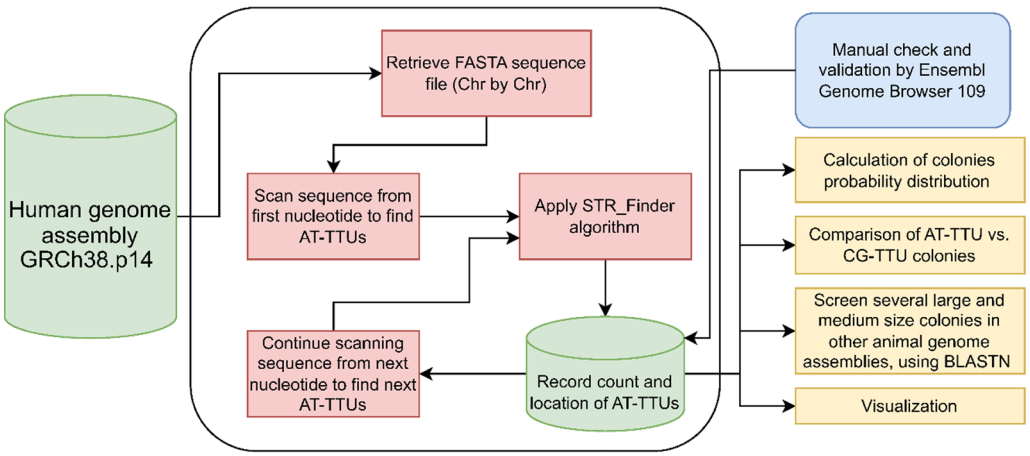

Fig. 1 Workflow diagram, outlining the various steps and algorithm developed in this research

these units across the entire genome. Through this verification process, we confirmed that the algorithm functioned as intended, and produced reliable results.

Comparison of the AT-TTU and CG-TTU colonies in the human genome

The AT-TTU colonies from the present study were compared to the CG-TTU colonies, yielded from our previous study (https://figshare.com/articles/dataset/All_possible_CG-rich_trinucleotides/23260562) [21].

The extraction algorithm

The developed Java program was used to extract all possible AT-TTUs, as follows: AATAAT, ATAATA, ATTATT, TTATTA, TATTAT, and TAATAA, from the human genome sequence. The program initiated its search from the first nucleotide of the genome, continuously scanning for the occurrence of AT-TTUs. It employed a window frame, consisting of 6 nucleotides. Upon discovering an AT-TTU, the program recorded the count and location of the occurrence. It then proceeded to search for new AT-TTUs, starting from the next nucleotide.

To validate the results, the final list of the identified AT-TTUs underwent manual evaluation, using the Ensembl genome browser 109 (https://asia.ensembl.org/index.html). The precise locations of the AT-TTUs were determined as follows: The output was organized and classified in an Excel file, where the start and end points of each AT-TTU were determined in the genome. By subtracting the start and end points of consecutive AT-TTUs, colonies were identified. If the resulting distance between consecutive AT-TTUs was less than 500 bp, these AT-TTUs were considered part of the same colony. Subsequently, a list of colonies, consisting of two or more AT-TTUs was compiled, the total count of colonies was determined, and the output was saved in a readily available format (https://doi.org/10.6084/m9.figshare.24202461.v1).

Screening several large and medium-size human colonies in five other species

Several of the large and medium-size colonies in human were screened in five other species, spanning primates and mouse, using the Genome Browser 109 (https://asia.ensembl.org/index.html) BLASTN program. This investigation also included checking of the flanking sequences of the AT-TTUs, to ensure specificity of the colonies in these species. The genome assemblies used were as follows: Chimpanzee: Pan_tro_3.0, Gorilla: gorGor4, Macaque:Mmul_10, Mouse lemur: Mmur_3.0, and Mouse: GRCm39.

Statistical analysis

The Poisson distribution was employed to determine the probability distribution of the AT-TTU colonies. This model assumed that the occurrence of the AT-TTUs was random and independent of each other, and that their distribution across the genome was relatively even. This assumption was subsequent to our observations of the ubiquitous occurrence of the sample colonies studied (Table 1). Therefore, the probability of occurrence of various size colonies was calculated by the Poisson density function, using the following formula:

$$\lambda = \frac{\textit{Colony Interval } (bp)^{*}\text{all possible AT} - \text{TTUs in the human genome}}{\text{Genome Size}(3\text{ gb})}$$

For example, the largest colony in human, C718, spanned 21,859 bp of genomic DNA. On the other hand, the total count of AT-TTUs in the human genome was about 10,330,879, resulting in $\lambda = 75.27$ for C718, meaning that based on the Poisson distribution, the average expected count of AT-TTUs in the 21,859 bp interval was 75.27. Table 1 presents values of λ for several colony sizes. The calculated probability value of the occurrence of these colonies was inherent zero.

Visualization

The six pure AT-TTUs were visualized as: **TTATTA**, **TATTAT**, **AATAAT**, **ATTATT**, **ATAATA**, and **TAATAA**. All the overlapping units were also highlighted, using various highlight and text colors.

Results

The majority of the AT-TTUs resided in colonies.

In total, 10,330,879 AT-TTUs were detected across the human genome, of which the majority (9,936,861) (96.18%) were arranged in 1,390,055 colonies (Fig. 2) (Suppl. 1). The AT-TTUs were spread across all chromosomes (Fig. 3).

AT-TTU colonies were significantly more abundant and larger than the CG-TTU colonies

In comparison to the CG-TTU colonies (https://figshare.com/articles/dataset/All_possible_CGrich_trinucleot

Table 1 Several large and medium-size AT-TTU colonies in human and their corresponding colonies in other primates

Colony formula[b]							
Colony size in human[a]	**Human**	**Chimpanzee**	**Gorilla**	**Macaque**	**Chr. No.**	**Location (Colony interval)**	**λ**
C718	[(ATT)2]32 [(ATA)2]5 [(AAT)2]11 [(TAA)2]2 [(TTA)2]318 [(TAT)2]350	[(ATT)2]21 [(ATA)2]5 [(AAT)2]3 [(TAA)2]2 [(TTA)2]262 [(TAT)2]288	[(ATT)2]3 [(ATA)2]2 [(TAA)2]1 [(TTA)2]4 [(TAT)2]7		11	11:114603789–114625648	75.27
C457	[(ATT)2]13 [(ATA)2]160 [(AAT)2]40 [(TAA)2]43 [(TTA)2]8 [(TAT)2]193				22	22:18728881–18733753	16.78
C350	[(ATT)2]12 [(ATA)2]165 [(AAT)2]15 [(TAA)2]11 [(TTA)2]11 [(TAT)2]136				11	11:96561723–96568076	21.88
C317	[(ATT)2]115 [(ATA)2]59 [(AAT)2]8 [(TAA)2]4 [(TTA)2]14 [(TAT)2]117	[(ATT)2]102 [(ATA)2]47 [(AAT)2]5 [(TAA)2]3 [(TTA)2]12 [(TAT)2]103			12	12:79456025–79468112	41.62
C287	[(ATT)2]11 [(ATA)2]112 [(AAT)2]30 [(TAA)2]13 [(TTA)2]16 [(TAT)2]105	[(ATT)2]2 [(ATA)2]17 [(AAT)2]1 [(TAA)2]1 [(TTA)2]2 [(TAT)2]24		[(ATT)2]2 [(ATA)2]3 [(AAT)2]2 [(TAA)2]1 [(TTA)2]2 [(TAT)2]6	X	X:143653179–143659751	22.63
C275	[(ATT)2]6 [(ATA)2]135 [(AAT)2]13 [(TAA)2]10 [(TTA)2]6 [(TAT)2]105				22	22:18891229–18896934	19.65
C267	[(ATT)2]4 [(ATA)2]151 [(AAT)2]4 [(TAA)2]4 [(TTA)2]5 [(TAT)2]99				2	2:16146128–16148756	9.05
C255	[(ATA)2]4 [(AAT)2]247 [(TAA)2]2 [(TTA)2]1 [(TAT)2]1	[(ATA)2]2 [(AAT)2]16 [(TAA)2]1 [(TTA)2]1 [(TAT)2]1			2	2:231822812–231850574	95.6

Table 1 (continued)

Colony formula[b]							
Colony size in human[a]	**Human**	**Chimpanzee**	**Gorilla**	**Macaque**	**Chr. No.**	**Location (Colony interval)**	**λ**
C212	[(ATA)2]101	[(ATT)2]20		[(ATT)2]1	X	X: 108772450–108776693	14.61
	[(AAT)2]4	[(ATA)2]95		[(ATA)2]17			
	[(TAA)2]6	[(AAT)2]20		[(AAT)2]1			
	[(TTA)2]2	[(TAA)2]16		[(TAA)2]3			
	[(TAT)2]99	[(TTA)2]8		[(TTA)2]1			
		[(TAT)2]88		[(TAT)2]20			
C200	[(ATT)2]4	[(ATT)2]2		[(ATA)2]80	7	7:37785667–37794029	28.8
	[(ATA)2]77	[(ATA)2]72		[(AAT)2]78			
	[(AAT)2]76	[(AAT)2]70		[(TAA)2]10			
	[(TAA)2]9	[(TAA)2]7		[(TTA)2]3			
	[(TTA)2]2	[(TTA)2]2		[(TAT)2]29			
	[(TAT)2]32	[(TAT)2]30					
C198	[(ATT)2]9				22	22:18235269–18238724	11.9
	[(ATA)2]62						
	[(AAT)2]14						
	[(TAA)2]15						
	[(TTA)2]5						
	[(TAT)2]93						
C195	[(ATT)2]12	[(ATT)2]10	[(ATT)2]6		10	10:67724852–67731164	21.74
	[(ATA)2]61	[(ATA)2]48	[(ATA)2]31				
	[(AAT)2]9	[(AAT)2]3	[(AAT)2]4				
	[(TAA)2]6	[(TAA)2]4	[(TAA)2]4				
	[(TTA)2]15	[(TTA)2]10	[(TTA)2]4				
	[(TAT)2]92	[(TAT)2]58	[(TAT)2]17				
C190	[(ATT)2]2				2	2:194462444–194465074	9.06
	[(ATA)2]38						
	[(AAT)2]5						
	[(TAA)2]6						
	[(TTA)2]3						
	[(TAT)2]136						
C184	[(ATT)2]34	[(ATT)2]23	[(ATT)2]19	[(ATT)2]6	6	6:77601972–77604184	7.62
	[(ATA)2]18	[(ATA)2]10	[(ATA)2]10	[(ATA)2]1			
	[(AAT)2]30	[(AAT)2]19	[(AAT)2]17	[(AAT)2]3			
	[(TAA)2]4	[(TAA)2]3	[(TAA)2]2	[(TAA)2]1			
	[(TTA)2]61	[(TTA)2]34	[(TTA)2]26	[(TTA)2]2			
	[(TAT)2]37	[(TAT)2]24	[(TAT)2]19	[(TAT)2]5			
C182	[(ATT)2]13	[(ATT)2]6	[(ATT)2]4	[(ATT)2]3	14	14:52527096–52531874	16.45
	[(ATA)2]51	[(ATA)2]65	[(ATA)2]31	[(ATA)2]11			
	[(AAT)2]17	[(AAT)2]12	[(AAT)2]2	[(AAT)2]2			
	[(TAA)2]16	[(TAA)2]11	[(TAA)2]2	[(TAA)2]3			
	[(TTA)2]12	[(TTA)2]9	[(TTA)2]6	[(TTA)2]7			
	[(TAT)2]73	[(TAT)2]38	[(TAT)2]22	[(TAT)2]15			

Table 1 (continued)

Colony formula[b]							
Colony size in human[a]	**Human**	**Chimpanzee**	**Gorilla**	**Macaque**	**Chr. No.**	**Location (Colony interval)**	**λ**
C175	[(ATT)2]5				7	7:54395223–54397576	8.1
	[(ATA)2]68						
	[(AAT)2]5						
	[(TAA)2]7						
	[(TTA)2]23						
	[(TAT)2]67						
C173	[(ATA)2]69				16	16:18565943–18570325	15.09
	[(AAT)2]64						
	[(TAA)2]18						
	[(TAT)2]22						
C161	[(ATT)2]68	[(ATT)2]65	[(ATT)2]68		3	3:8693642–8702099	29.12
	[(ATA)2]17	[(ATA)2]17	[(ATA)2]17				
	[(AAT)2]2	[(AAT)2]2	[(AAT)2]3				
	[(TAA)2]2	[(TAA)2]2	[(TAA)2]2				
	[(TTA)2]8	[(TTA)2]8	[(TTA)2]4				
	[(TAT)2]64	[(TAT)2]64	[(TAT)2]65				

[a] Colony size, chromosomal location, colony interval, and λ are based on the human genome, as reference. The corresponding colonies in other species were identified, using BLASTN. Instances in which the colonies were partially sequenced (such as C287, C212, and C200 in gorilla), or lacked the corresponding colony in a species were left blank. None of the colonies in this table were detected in mouse lemur or mouse

[b] Formulas represent absolute count of units, regardless of being pure or overlapping. The Poisson probability of the colonies was inherent zero

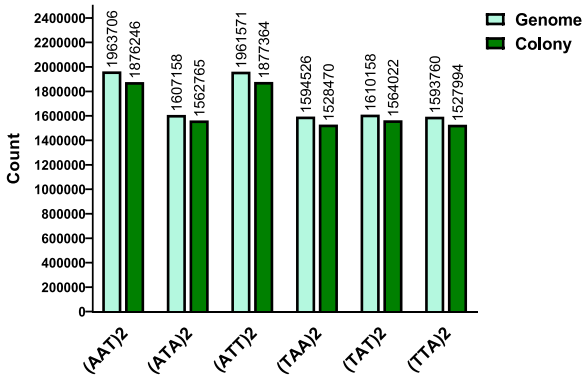

Fig. 2 Genome-wide count of AT-TTUs in human. The majority of the AT-TTUs were arranged in colonies. Absolute counts are depicted

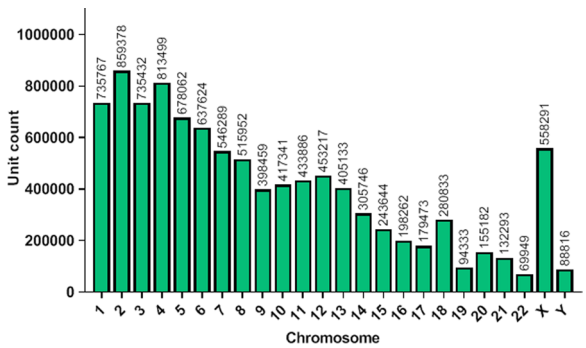

Fig. 3 Count of AT-TTUs across human chromosomes. Chromosome-by-chromosome absolute count of all possible AT-TTUs is depicted

ides/23260562), the colonies formed by AT-TTUs were significantly more abundant (Fig. 4). Large intervals of chromosomes were occupied by colony intervals in many chromosomes, for example in chromosome 4. Furthermore, the pattern of distribution of the AT-TTU colonies across human chromosomes was significantly different from the CG-TTU colonies. For example, whereas chromosome 1 had the highest percentage of CG-TTU colonies [21], AT-TTU colonies reached highest percentage on chromosome 4. Chromosome X was also enriched by AT-TTU colonies.

Several of the large and medium-size AT-TTU colonies coincided with extensive dynamicity in great apes

Several of the large and medium-size AT-TTU colonies in human were also detected in other great apes

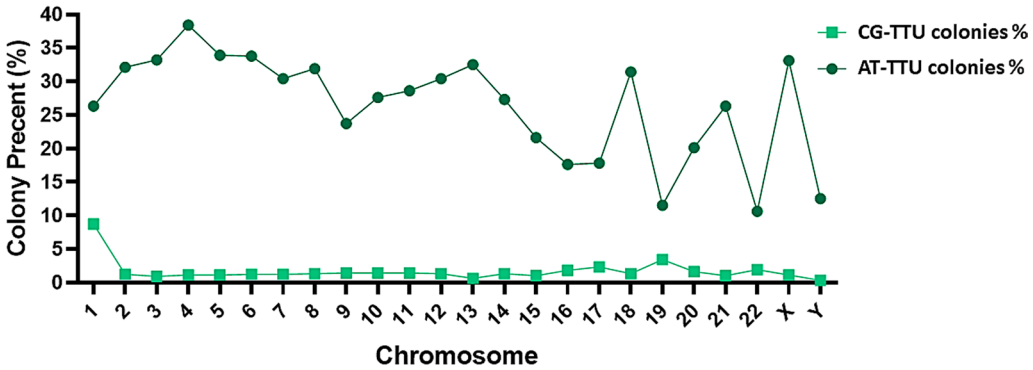

Fig. 4 Normalized distribution of AT-TTU vs. CG-TTU colonies across human chromosomes. The AT-TTU colonies were significanlty more abundant than the CG-TTU colonies, and occupied significant intervals of several chromosomes (maximally in chromosome 4). Colony percentage (Y-axis) depicts the percentage of each chromosome that is occupied by the AT-TTU colonies. The CG-TTU data were extracted from the following link: https://figshare.com/articles/dataset/All_possible_CG-rich_trinucleotides/23260562

(Table 1). Exceedingly dynamic events were detected across these colonies, affecting the AT-TTUs and the flanking sequences to the units. Across the colonies, the AT-TTUs were either pure or overlaps of two or more pure units.

The largest AT-TTU colony in human was a compound colony of 718 units (C718), located on chromosome 11, which was detected with exceeding dynamicity in human and chimpanzee, and at a far lesser extent in gorilla. This colony reached maximum complexity and size in human (Fig. 5). The absolute count of the AT-TTUs and the distribution of the units in the pure and overlapping compartments were exceedingly dynamic across these species, adding multiple layers of complexity of the events, and leading to massively divergent compositions.

Most of the units in C718 and its orthologous colonies were in the overlapping compartment (Figs. 5 and 6A). The immediate flanking sequences of the overlapping units conformed to the flanking sequences of the involved pure units, and were significantly dynamic with respect to mutations (Fig. 5B).

Models proposed for the evolution of pure and overlapping units

The pure units were the inverted or palindromic sequences of one another, and probably resulted in DNA breakage and recombination events inherent to inverted and palindromic sequences, for example, two pure units of TTATTA and ATTATT (inversion), and TTATTA and TAATAA (palindrome).

Overlapping units were a consequence of unequal crossovers among the pure units. For example, in C718, the most prevalent overlapping unit, TTATTAT, was the consequence of unequal crossovers between pure units, TTATTA and TATTAT (Fig. 6A). In another example in C718, the overlapping unit, AATAATTATTAT, was the consequence of several unequal crossovers across units (Fig. 6A). It is conceivable that reverse processes leading to the overlapping units resulted in the re-emergence of the pure units.

The flanking sequences of the units were also highly dynamic (Fig. 6B), signifying the occurrence of crossovers at the sites of the AT-TTUs, and coupled breakage and repair at, and around these sites.

Coincidence of some of the colonies beyond great apes

Several colonies, such as C212, C200, and C184 coincided beyond great apes, and included macaque (Table 1). As an example, in C184, the colonies were shared dynamically in human, chimpanzee, gorilla, and macaque, and there was a directional incremented trend of complexity of the events and units in human (Fig. 7). Pure and overlapping units were also detected across this colony in human and other primates. For example, TATTATTA, was the consequence of unequal crossovers between TAT TAT, ATTATT, and TTATTA pure units.

Some colonies were detected in human and not the other five species studied

We also detected colonies that were found in human only (Table 1), examples of which are visualized for C457 (Fig. 8A) and C190 (Fig. 8B). Consecutive pure units recombining with each other, or pure and overlapping units recombining with each other were detected in these colonies.

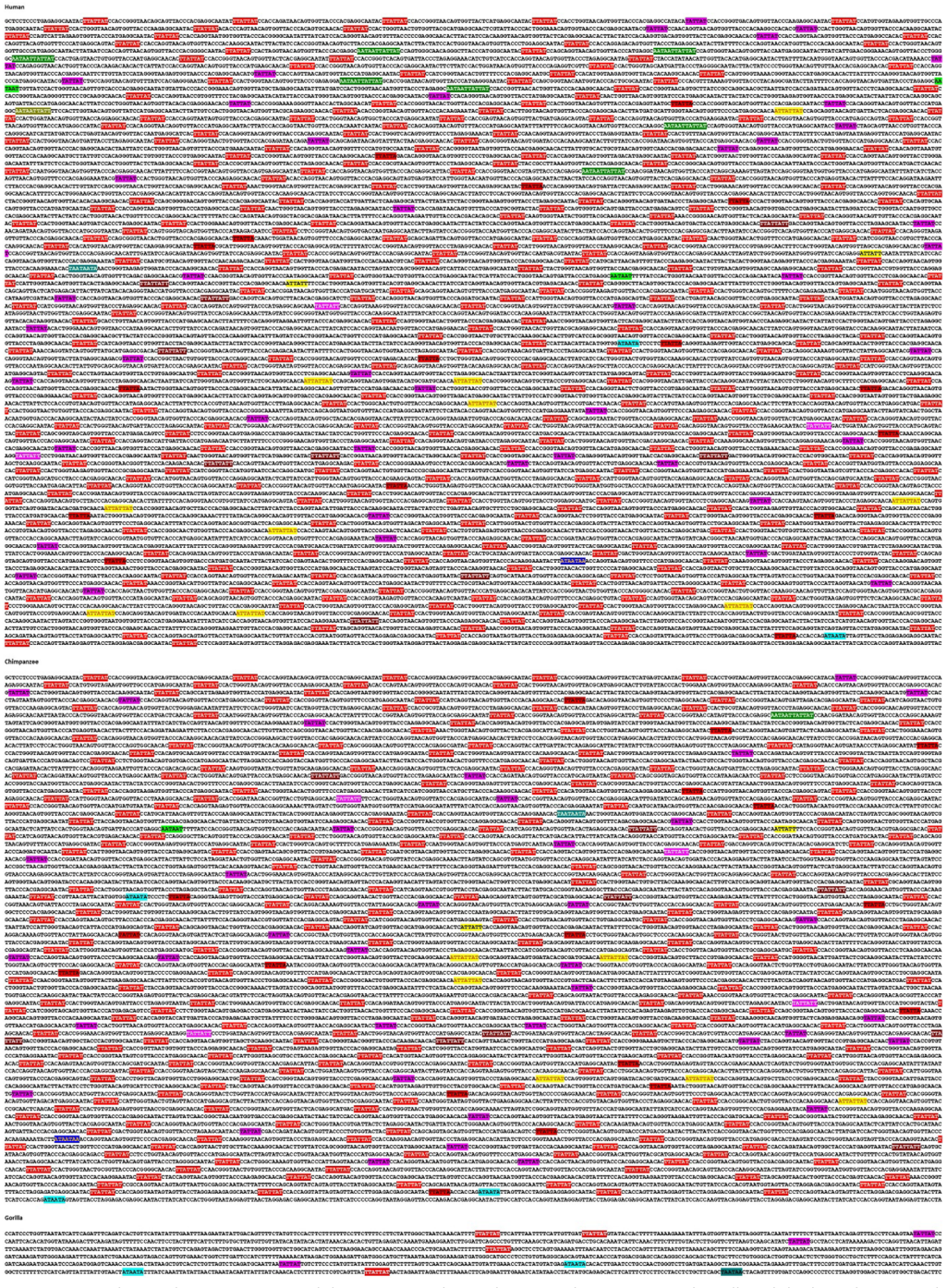

Fig. 5 The largest AT-TTU colony in human (C718) and the corresponding colonies in chimpanzee and gorilla. While the colony was shared across these apes, we detected dynamic differences and species-specific formulas and compositions of the AT-TTUs. Pure units and overlapping units of the pure units were detectable, signifying sites of unequal cross-over at the units. The colony reached maximum complexity and size in human

AT-TTUs are a mechanism for the emergence of A and T short tandem repeats (STRs)

The AT-TTUs and coupled unequal crossovers and recombination at these sites result in the emergence of STRs (repeats of ≥ 3). For example, in C184, the (TTA)3 STR could be a consequence of unequal crossovers through various paths (Fig. 9A and B). In other examples, in C457 and C190, unequal crossovers gave rise to

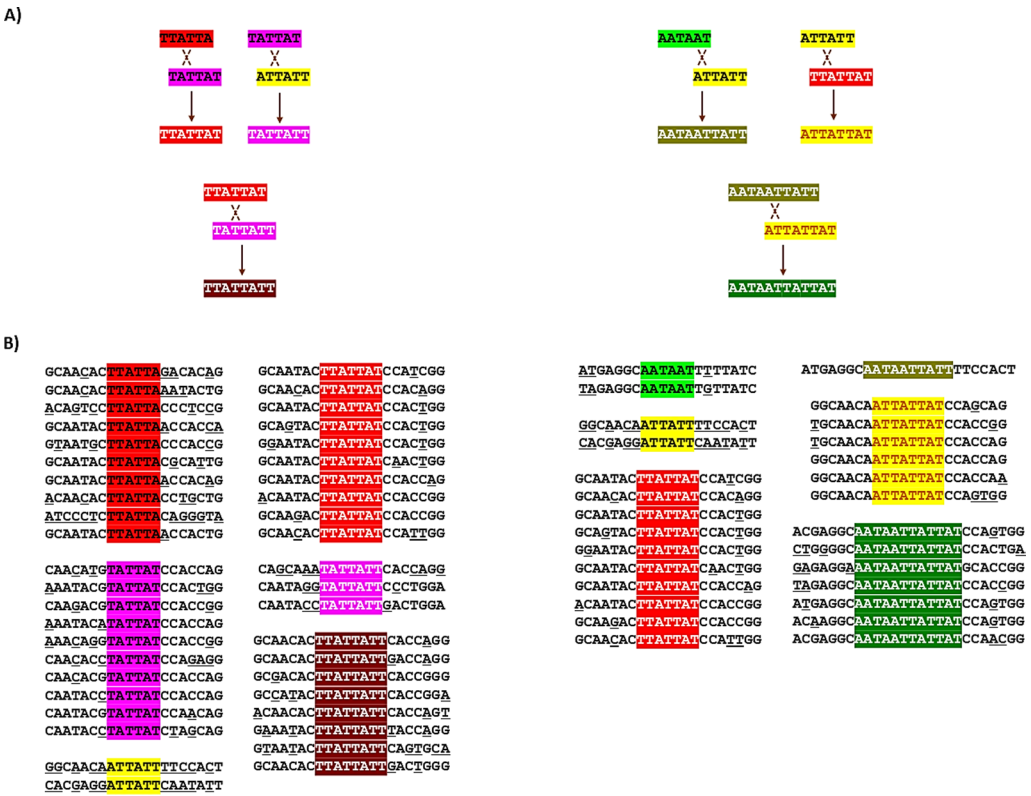

Fig. 6 Emergence of overlapping units from pure units. Emergence of the most prevalent overlapping unit in C718, and other overlapping units in this colony (**A**). For simplicity, only the alleles involved in the process of gaining overlapping units are depicted. A sample of the flanking sequences to each unit is depicted (**B**). For the units that were highly prevalent, only 10 sequences were randomly selected from the human C718 colony. The flanking sequences of the overlapping units conformed to the flanking sequences of the involved pure units, and were significantly dynamic with respect to mutations. Underlines represent probable mutations (the least frequent substitutions in a given nucleotide position are underlined). The high density of flanking mutations is an expected consequence of the unequal crossovers at the units and breakage/repair mechanisms at, and around these sites. The models represent only a sample of the dynamicity at the units and their flanking sequences

overlapping units for the emergence of several (ATA)3 STRs (Fig. 8C, D, and E). We detected the pure units and intermediate overlapping units necessary for the emergence of a given STR, in the same (or orthologous) colonies that the STR was detected.

AT-TTUs may regulate transposable elements (TEs)

We observed that some of the colonies, such as C718, were surrounded by various classes of TEs, such as short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and long terminal repeats (LTRs) (https://genome.ucsc.edu/), whereas within the colony interval was mainly devoid of these elements. This property was observed in human, chimpanzee, and gorilla, for C718 (Fig. 10). This colony may function as a potential *cis* inhibitor of TEs in the human genome.

Discussion

The bulk of literature is dominated by reports of the preference of CG- over AT-rich sequences at the recombination hotspots [15, 22–27]. Limited reports of the involvement of AT-rich sequences in recombination and consequent translocations primarily concern AT-rich palindromic or inverted sequences. These events are mainly involved in chromosomal translocations and deletions, for example in chromosomes 11, 17, and 22 [28–30].

The algorithm developed in the present study aimed at including palindromes and inversions in the context

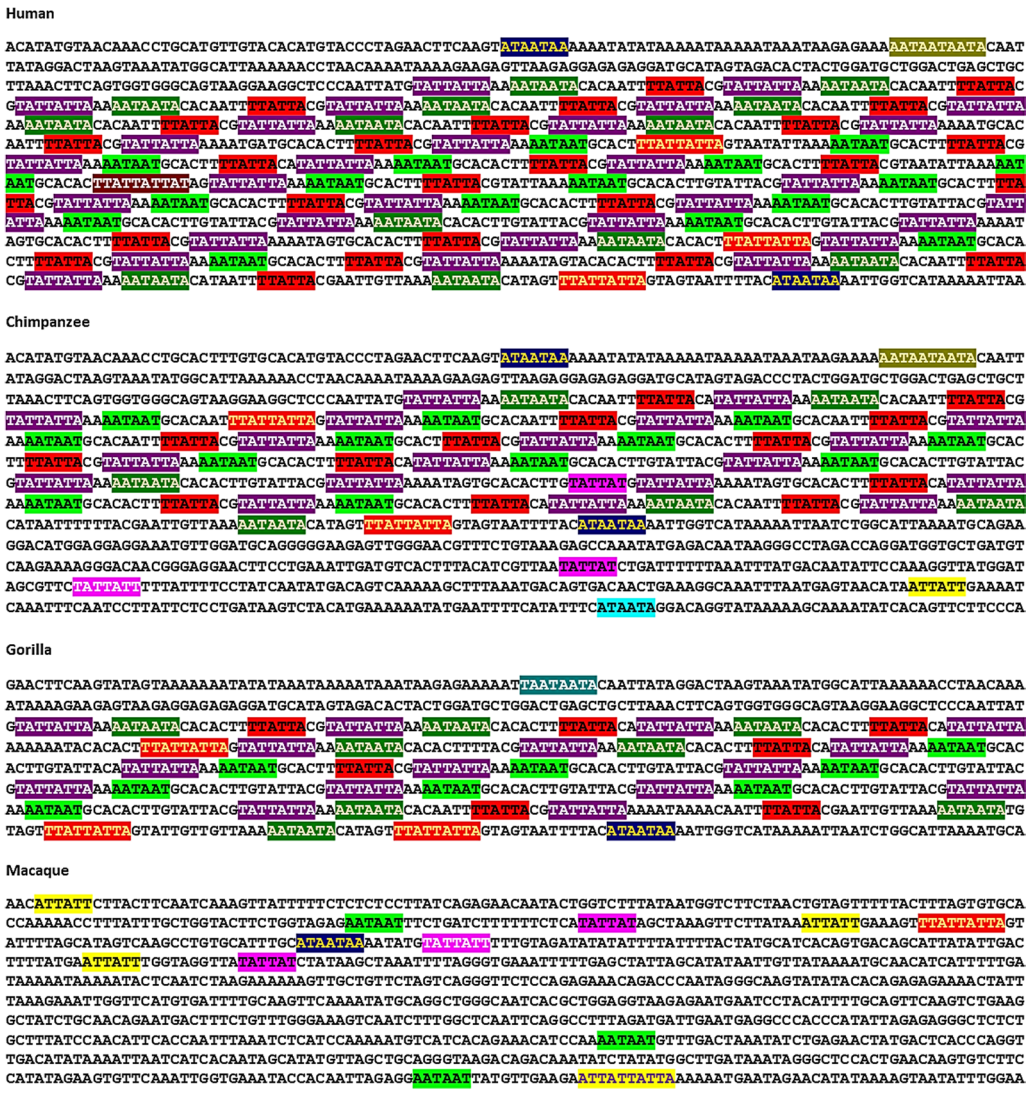

Fig. 7 Example colony shared across great apes and macaque (C184). High dynamicity encompassed the AT-TTUs, as well as the flanking sequences to each unit. The colony reached maximum complexity and size in human. It is conceivable that the pure units, AATAAT and TTATTA , emerged from unequal crossovers between sister chromatids and non-sister homologous chromosomes. It can also be predicted that AATAAT emerged before TTATTA, as the former was detectable as distantly as in macaque, whereas TTATTA was not detected in this species. Overlapping units of these units and other pure units later emerged in gorilla, chimpanzee, and human

of AT-TTU pure units, which led to the identification of a phenomenon, whereby AT-TTUs colonized across the genome with exceeding significance, based on Poisson distribution. In fact, the majority of AT-TTUs resided in colonies, and these colonies spanned significant intervals of several chromosomes. Remarkably, chromosome X was also enriched by colonies.

The AT-TTU colonies were significantly larger and more complex than the CG-TTU colonies that we reported previously [20, 21]. These findings support a more significant role of AT-rich sequences in comparison to CG-rich sequences, as crossover and recombination hotspots. The AT-TTU crossover hotspots are ubiquitous, whereas the most refined maps of recombination identified to date are on the cM scales [31, 32].

The presence of pure units and overlapping units of the pure units, signify that the main reason for the hotspot events in the colonies is the AT-TTUs. The inversions and palindromes as a result of the pure and overlapping units increase the rate of various genetic rearrangement events and recombination across the colonies. Palindromes and inversions are known to be recombinogenic in the genomes, and a risk to instability [33, 34].

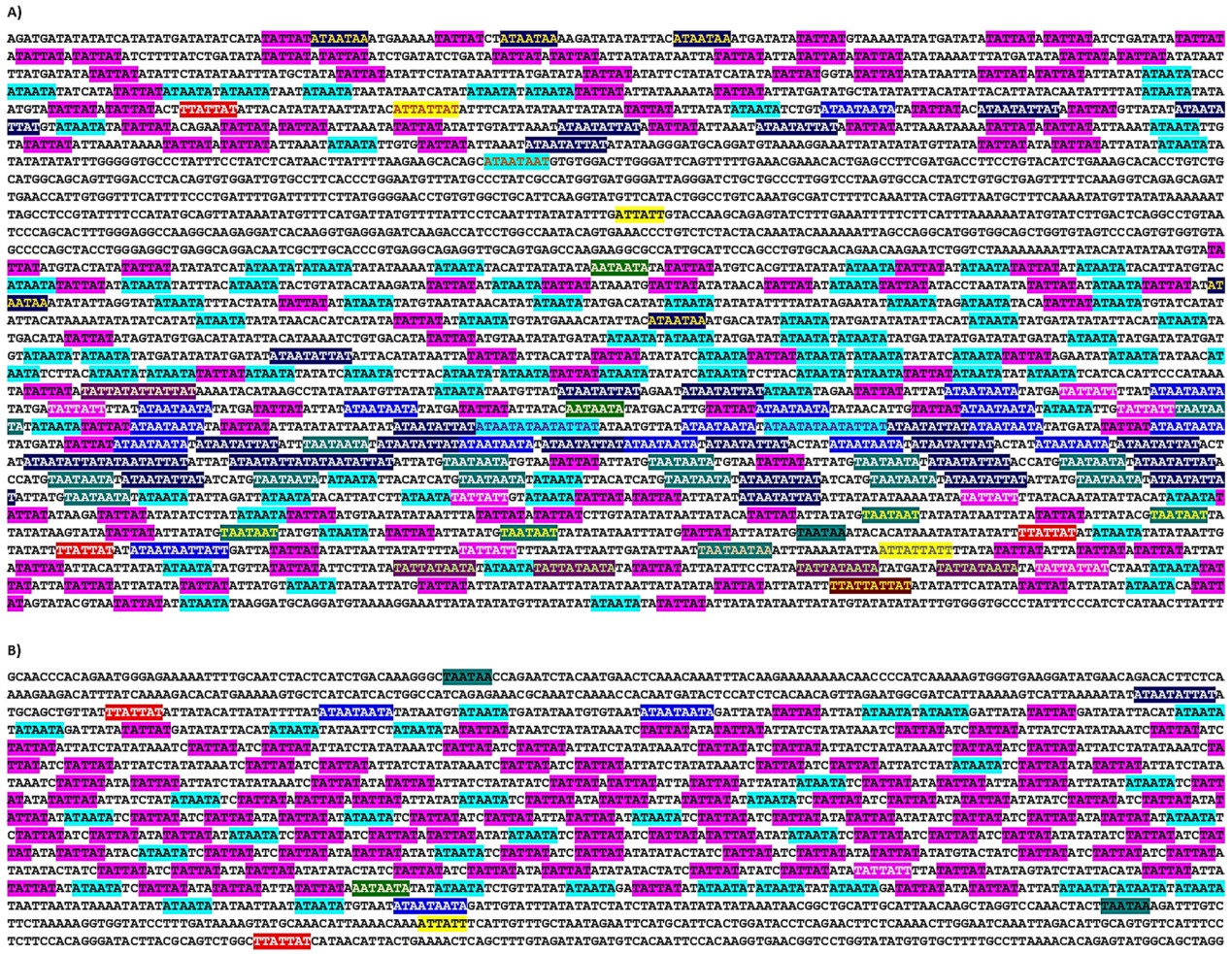

Fig. 8 Example colonies that were detected in human and not the other five species. These colonies were denser than the non-specific colonies. **A** C457 was the largest colony in the human-only category. The high intensity of unequal crossovers in C457 resulted in recombination of numerous pure and overlapping units in some regions, e.g., consecutive blue, purple, and navy units. Inversions and palindromes were readily detectable. For example, the ATAATATATTAT palindrome was the consequence of the recombination of two pure units, ATAATA and TATTAT. **B** C190 exemplifies a medium-size colony of mainly pure units. This colony was also highly dynamic with respect to the intensity of AT-TTU recombination

The flanking sequences of the AT-TTUs were also extensively dynamic. The very high dynamicity of the flanking sequences was in line with the previous reports that flanking sequences to the recombination sites are prone to mutations [35].

Some of the identified colonies, which were further studied in several other primates, were shared in these primates with exceeding dynamicity of the events. These findings challenge the literature on the rarity of shared recombination hotspots between human and closely related species [36–39]. An isolate report of shared hotspot loci between human and chimpanzee was at β-globin and HLA regions on chromosome 21, which was based on high Bayes factors of shared hotspots at locations within both regions [40]. Our data extend crossover and recombination hotspot sharedness across primates, and envision a new perspective with respect to the magnitude of these events across genomes.

It is reasonable to consider AT-TTUs a novel genomic entity, as although they are repeats, they do not conform to the conventional definition of repetitive DNA sequences [41]. It is also expected that novel proteins are recruited to these loci, as neither the well-characterized recombination hotspot 13-mer, nor the degenerate sequences of this sequence conform to the identified AT-TTUs and colonies, and the extent of the events occurring across these colonies.

It is possible that some of the AT-TTU colonies are a result of "non-crossover" recombination, which includes exchange of DNA fragments, without exchanging the

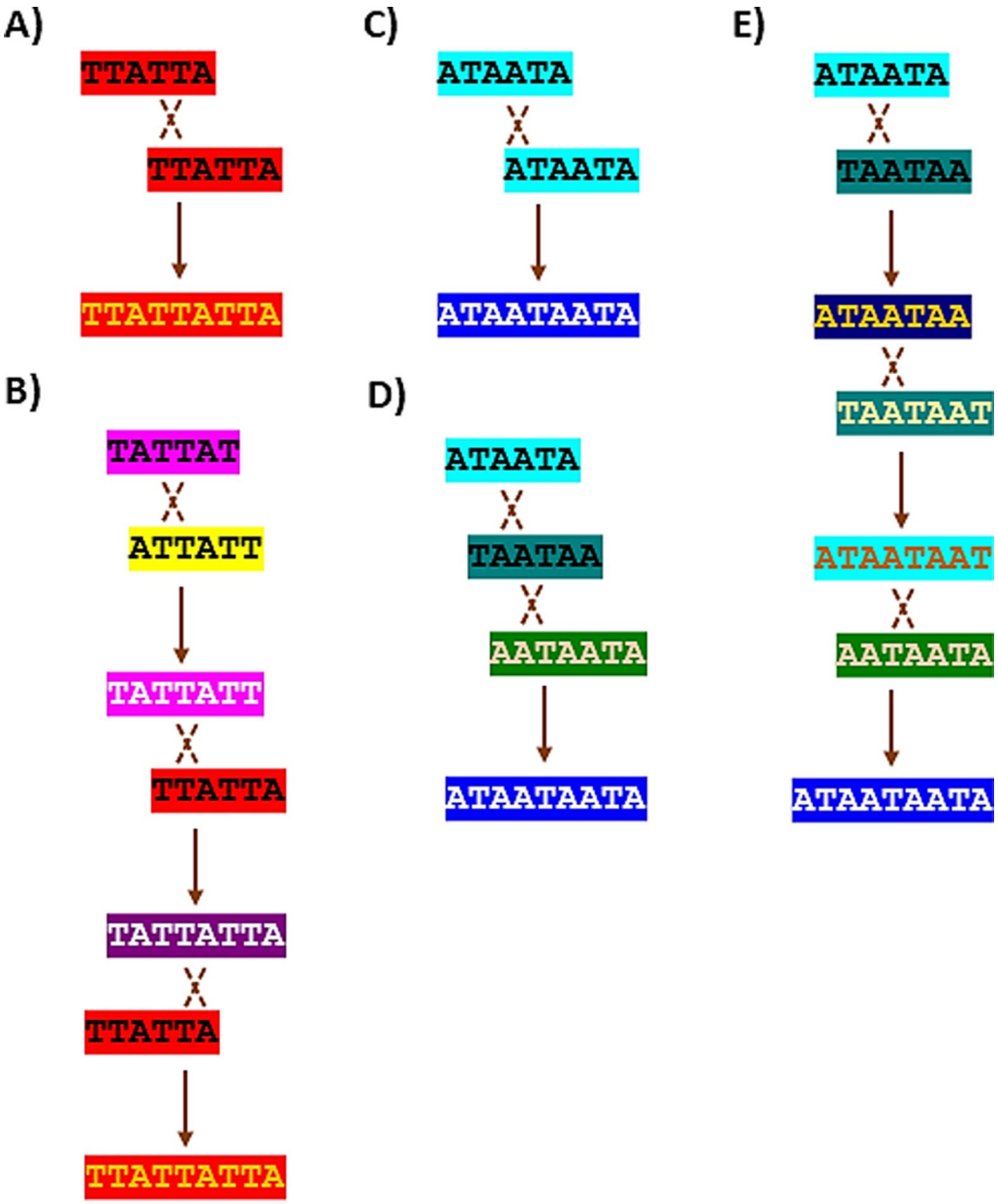

Fig. 9 AT-TTUs are a novel mechanism for the emergence of STRs. For simplicity, only the alleles involved in the process of STR emergence are depicted. Models of the birth and maturation of (TTA)3 STRs in C184 (**A**) and (**B**), (ATA)3 STRs in C457 (**C**), (**D**), and (**E**) and (ATA)3 STRs in C190 (**C**) and (**D**)

flanking chromosome arm. In fact, the majority of recombination interactions in meiosis are of the non-crossover type, as opposed to crossovers, which include the flanking chromosomal arm as well. Similar to crossovers, knowledge on the sites and biological implications of non-crossovers are also limited (if not less) at this time, and they are more difficult to detect. Evidence indicates that there probably is no non-crossover-specific pathway, and that restoration of intermediate events in a single pairing/recombination pathway promotes synaptonemal complex formation [42]. PRDM9 recruitment, CG-bias at the sites of recombination, and nearby conversions are also inherent to non-crossovers known to date [22, 43]. The AT-TTUs identified here, unveil recombination hotspots and evolutionary implications, which may be of relevance in both crossover and non-crossover contexts. Throughout this paper, we used the term "crossover" as its general application (see Glossary). That term was not

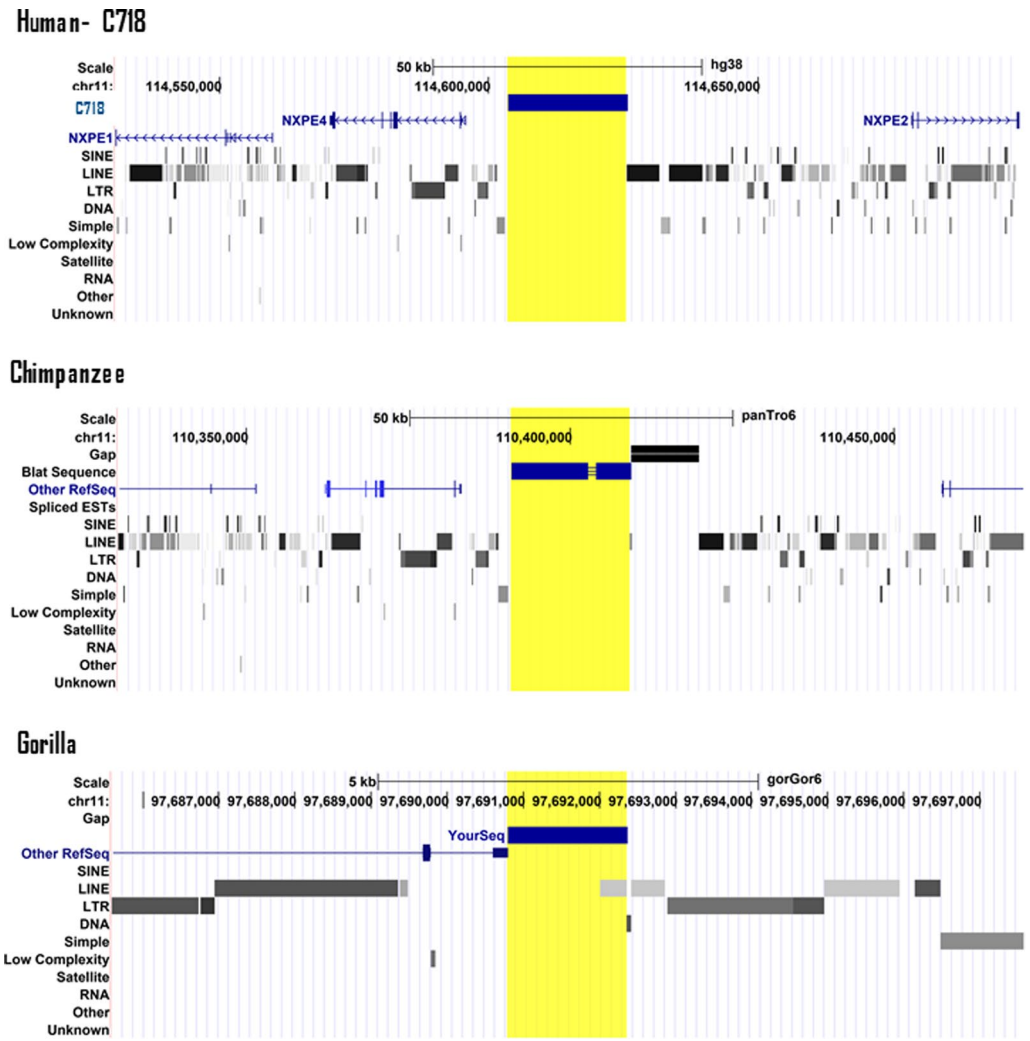

Fig. 10 Potential inhibitory effect of C718 on surrounding TEs. While this colony is surrounded by various TEs, such as SINEs, LINEs, and LTRs, the colony interval itself is mainly devoid of these elements (Blat Search in https://genome.ucsc.edu/). C718 in human and the orthologous colonies in chimpanzee and gorilla are yellow-highlighted

intended to differentiate between crossover and non-crossover events defined above.

In comparison to CG-TTU colonies, the AT-TTU colonies (at least the colonies that were further analyzed in additional primates and mouse), were mainly more complex in human, at a directional trend. Furthermore, the rate of detecting these colonies in human only, was higher than the CG-TTU colonies [21]. One explanation may be that the mechanisms involved in the development of the AT-TTU colonies evolved more recently than the CG-TTU colonies. This is also supported by our observations that the sample colonies studied in Table 1 were not detected in mouse lemur and mouse, whereas in the instance of CG-TTU colonies, several of the colonies were identified in these species [21]. However, it should be noted that more comprehensive evolutionary studies are warranted to draw solid conclusions on the evolutionary time-scale of the AT- and CG-TTU colonies.

It is estimated that approximately 50% of the human genome contains repeat elements [44]. These elements are classified into different classes, including STRs, LINEs, SINEs, LTRs, minisatellite and satellite repeats, RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA), other repeats e.g., class RC (Rolling Circle), and Unknown [45]. Similar to CG-TTUs [21], AT-TTUs and the crossovers coupled with these units are a novel mechanism for the emergence of STRs. This follows from our observations that all the pure and intermediate overlapping units necessary for the birth and maturation of a given STR were detectable in the same

(or orthologous) colonies. STRs are being increasingly linked to significant functions of evolutionary, biological, and pathological consequences [46–50]. The colonies may also be coupled with the inhibition of TEs, such as SINEs, LINEs, and LTRs. TEs contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes [51]. It is, therefore, conceivable that the interaction between TEs and the identified colonies will eventually shape the genome structure and functionality.

Considering that large intervals of chromosomes are occupied by the AT-TTU colonies and the recombination events coupled with these colonies, it is conceivable that these colonies link to genome size regulation. This concept is in line with several reports, for example, the driving force of recombination on vertebrate genome size evolution [52, 53], and the chromosome size effect on sequence divergence among species through the interplay of recombination and selection [54].

Taken together, in view of the events associated with the identified AT-TTUs, their abundance and ubiquity throughout the genomes studied, and exceedingly significant colonization based on Poisson distribution, we predict that these findings are the tip of the iceberg, various aspects of which are yet to be explored in the future studies.

Conclusion

Our findings unveil massive AT-TTU crossover and recombination hotspots across the human genome, and signify preference of AT- over CG-rich sequences at the crossover and recombination hotspots. These recombination hotspots are conserved, yet with extensive dynamicity, at least across great apes and Old-World monkeys.

Abbreviations

AT-TTU	A and T trinucleotide two-repeat unit
C	Colony
CG-TTU	C and G trinucleotide two-repeat unit
cM	CentiMorgan
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
SINE	Short interspersed nuclear element
STR	Short tandem repeat
TE	Transposable element

Glossary

Unit	Two-repeats of any A and T trinucleotides. For example, TTATTA is a two-repeat unit of the TTA trinucleotide.
Colony	A group of units, in which the distance between two consecutive units was < 500 bp. Throughout the text, specific colonies are identified by adding the "C" prefix, where necessary. The designation of colonies is based on their size in the human genome. For example, C718 is a colony of 718 units in human.
Compound Colony	A colony that consists of more than one type of two-repeat units of AT trinucleotides. For example, a compound colony could include (TTA)2 and (TAT)2 units.
Pure unit	A unit that consists of only one type of A and T trinucleotide, for example, TTATTA.
Overlapping unit	A unit that consists of two or more pure units that overlap. For example, the sequence "TTATTATT" consists of three pure units of TTATTA, TATTAT, and ATTATT, which overlap with each other.
Absolute count	Count of units regardless of being pure or overlapping.
Crossover	The exchange of DNA between paired homologous chromosomes (one from each parent) that occurs during the development of egg and sperm cells (meiosis).
Unequal crossover	Unequal crossing-over, also referred to as illegitimate recombination, refers to crossover events that occur between nonequivalent sequences.
Non-crossover	Recombination interactions, which include DNA fragments, without exchange of flanking chromosome arms.
Recombination hotspot	A genomic region (typically in ~ kb ranges) that experience intensely high levels of Recombination compared to the genomic background.
Repeat of ≥ 3A	A genomic region (typically in ~kb ranges) that experience intensely high levels of Recombination compared to the genomic background.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13062-024-00508-8.

Additional file1

Acknowledgements

Not applicable.

Author contributions

Conceptualisation: MO; Methodology: MA, MAMA; Investigation: MA, SKh, SA, SV, HB, NT; Visualization: MA, SA, SKh; Project administration: MO, AD, HRKh; Supervision: MO; Writing – original draft: MO, MA; Writing – review & editing: MO, MA.

Funding

Not applicable.

Availability of data and materials

Raw data for AT-TTUs are available at the following link: https://figshare.com/articles/dataset/AT-rich_trinucleotides/24202461. Raw data for CG-TTUs are available at the following link: https://figshare.com/articles/dataset/All_possible_CG-rich_trinucleotides/23260562.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 May 2024 Accepted: 29 July 2024
Published online: 21 August 2024

References

1. Ortiz-Barrientos D, Engelstadter J, Rieseberg LH. Recombination rate evolution and the origin of species. Trends Ecol Evol. 2016;31(3):226–36.
2. Paigen K, Petkov P. Mammalian recombination hot spots: properties, control and evolution. Nat Rev Genet. 2010;11(3):221–33.
3. Wall JD, Stevison LS. Detecting recombination hotspots from patterns of linkage disequilibrium. G3. 2016;6(8):2265–71.
4. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics. 2003;165(4):2213–33.
5. Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science. 2010;327(5967):876–9.
6. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2010;327(5967):836–40.
7. Parvanov ED, Petkov PM, Paigen K. Prdm9 controls activation of mammalian recombination hotspots. Science. 2010;327(5967):835.
8. Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet. 2008;40(9):1124–9.
9. Berg IL, Neumann R, Lam KW, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nat Genet. 2010;42(10):859–63.
10. Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ. Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. Proc Natl Acad Sci USA. 2011;108(30):12378–83.
11. Ubeda F, Fyon F, Burger R. The Recombination Hotspot Paradox: co-evolution between PRDM9 and its target sites. Theor Popul Biol. 2023;153:69–90.
12. Lartillot N. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. Mol Biol Evol. 2013;30(3):489–502.
13. Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJ. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol Biol Evol. 2013;30(9):2134–44.
14. Montoya-Burgos JI, Boursot P, Galtier N. Recombination explains isochores in mammalian genomes. Trends Genet. 2003;19(3):128–30.
15. Fullerton SM, Bernardo Carvalho A, Clark AG. Local rates of recombination are positively correlated with GC content in the human genome. Mol Biol Evol. 2001;18(6):1139–42.
16. Halo JV, Pendleton AL, Shen F, Doucet AJ, Derrien T, Hitte C, Kirby LE, Myers B, Sliwerska E, Emery S, Moran JV, Boyko AR, Kidd JM. Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes. Proc Natl Acad Sci USA. 2021;118(11):e2016274118.
17. Dutta R, Saha-Mandal A, Cheng X, Qiu S, Serpen J, Fedorova L, Fedorov A. 1000 human genomes carry widespread signatures of GC biased gene conversion. BMC Genomics. 2018;19(1):256.
18. Lachance J, Tishkoff SA. Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. Am J Hum Genet. 2014;95(4):408–20.
19. Borde V, Wu TC, Lichten M. Use of a recombination reporter insert to define meiotic recombination domains on chromosome III of *Saccharomyces cerevisiae*. Mol Cell Biol. 1999;19(7):4832–42.
20. Arabfard M, Tajeddin N, Alizadeh S, Salesi M, Bayat H, Khorram Khorshid HR, Khamse S, Delbari A, Ohadi M. Dyads of GGC and GCC form hotspot colonies that coincide with the evolution of human and other great apes. BMC Genomic Data. 2024;25(1):21. https://doi.org/10.1186/s12863-024-01207-z.
21. Ohadi M, Tajeddin N, Arabfard M, Alizadeh S, Bayat H, Moghadam MG, Khamse S, Salesi M, Maddi AMA, Delbari A. CG-rich trinucleotide two-repeats signify novel recombination hotspots conserved across primates and mouse. 2024.
22. Odenthal-Hesse L, Berg IL, Veselis A, Jeffreys AJ, May CA. Transmission distortion affecting human noncrossover but not crossover recombination: a hidden source of meiotic drive. PLoS Genet. 2014;10(2):e1004106.
23. Jeffreys AJ, Murray J, Neumann R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol Cell. 1998;2(2):267–73.
24. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 2004;14(4):528–38.
25. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 2009;10:285–311.
26. Marsolier-Kergoat MC, Yeramian E. GC content and recombination: reassessing the causal effects for the *Saccharomyces cerevisiae* genome. Genetics. 2009;183(1):31–8.
27. Charlesworth D, Zhang Y, Bergero R, Graham C, Gardner J, Yong L. Using GC content to compare recombination patterns on the sex chromosomes and autosomes of the guppy, *Poecilia reticulata*, and its close outgroup species. Mol Biol Evol. 2020;37(12):3550–62.
28. Kurahashi H, Inagaki H, Hosoba E, Kato T, Ohye T, Kogo H, Emanuel BS. Molecular cloning of a translocation breakpoint hotspot in 22q11. Genome Res. 2007;17(4):461–9.
29. Bi W, Park SS, Shaw CJ, Withers MA, Patel PI, Lupski JR. Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. Am J Hum Genet. 2003;73(6):1302–15.
30. Edelmann L, Spiteri E, Koren K, Pulijaal V, Bialer MG, Shanske A, Goldberg R, Morrow BE. AT-rich palindromes mediate the constitutional t(11;22) translocation. Am J Hum Genet. 2001;68(1):1–13.
31. Spence JP, Song YS. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. Sci Adv. 2019;5(10):eaaw9206.
32. Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. DNA recombination. Recombination initiation maps of individual human genomes. Science. 2014;346(6211):1256442.
33. Svetec Miklenic M, Svetec IK. Palindromes in DNA-A risk for genome stability and implications in cancer. Int J Mol Sci. 2021;22(6):2840.
34. Flores M, Morales L, Gonzaga-Jauregui C, Dominguez-Vidana R, Zepeda C, Yanez O, Gutierrez M, Lemus T, Valle D, Avila MC, Blanco D, Medina-Ruiz S, Meza K, Ayala E, Garcia D, Bustos P, Gonzalez V, Girard L, Tusie-Luna T, Davila G, Palacios R. Recurrent DNA inversion rearrangements in the human genome. Proc Natl Acad Sci USA. 2007;104(15):6099–106.
35. Kiktev DA, Sheng Z, Lobachev KS, Petes TD. GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 2018;115(30):E7109–18.
36. Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, Great Ape Genome P, Bustamante CD, Hammer MF, Wall JD. The time scale of recombination rate evolution in great apes. Mol Biol Evol. 2016;33(4):928–45.
37. Lesecque Y, Glemin S, Lartillot N, Mouchiroud D, Duret L. The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes. PLoS Genet. 2014;10(11):e1004790.
38. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D. Comparison of fine-scale recombination rates in humans and chimpanzees. Science. 2005;308(5718):107–11.
39. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Paabo S. Fine-scale recombination patterns differ between chimpanzees and humans. Nat Genet. 2005;37(4):429–34.
40. Wang Y, Rannala B. Bayesian inference of shared recombination hotspots between humans and chimpanzees. Genetics. 2014;198(4):1621–8.
41. Lower SE, Dion-Côté AM, Clark AG, Barbash DA. Special issue: Repetitive DNA sequences. Genes. 2019;10(11):896.
42. Storlazzi A, Xu L, Cao L, Kleckner N. Crossover and noncrossover recombination during meiosis: timing and pathway relationships. Proc Natl Acad Sci. 1995;92(18):8512–6.
43. Li R, Bitoun E, Altemose N, Davies RW, Davies B, Myers SR. A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. Nat Commun. 2019;10(1):3900.
44. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
45. Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, Gao X. Repetitive DNA sequence detection and its role in the human genome. Commun Biol. 2023;6(1):954.
46. Maddi AMA, Kavousi K, Arabfard M, Ohadi H, Ohadi M. Tandem repeats ubiquitously flank and contribute to translation initiation sites. BMC Genomic Data. 2022;23(1):59.

47. Arabfard M, Salesi M, Nourian YH, Arabipour I, Maddi AA, Kavousi K, Ohadi M. Global abundance of short tandem repeats is non-random in rodents and primates. BMC Genomic Data. 2022;23(1):77.
48. Horton CA, Alexandari AM, Hayes MGB, Marklund E, Schaepe JM, Aditham AK, Shah N, Suzuki PH, Shrikumar A, Afek A, Greenleaf WJ, Gordan R, Zeitlinger J, Kundaje A, Fordyce PM. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. Science. 2023;381(6664):eadd1250.
49. Tajeddin N, Arabfard M, Alizadeh S, Salesi M, Khamse S, Delbari A, Ohadi M. Novel islands of GGC and GCC repeats coincide with human evolution. Gene. 2021;902:148194. https://doi.org/10.1016/j.gene.2024.148194.
50. Alizadeh S, Khamse S, Tajeddin N, Khorram Khorshid HR, Delbari A, Ohadi M. A GCC repeat in RAB26 undergoes natural selection in human and harbors divergent genotypes in late-onset Alzheimer's disease. Gene. 2024;893:147968.
51. Diehl AG, Ouyang N, Boyle AP. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. Nat Commun. 2020;11(1):1796.
52. Nam K, Ellegren H. Recombination drives vertebrate genome contraction. PLoS Genet. 2012;8(5):e1002680.
53. Ross-Ibarra J. Genome size and recombination in angiosperms: a second look. J Evol Biol. 2007;20(2):800–6.
54. Tigano A, Khan R, Omer AD, Weisz D, Dudchenko O, Multani AS, Pathak S, Behringer RR, Aiden EL, Fisher H. Chromosome size affects sequence divergence between species through the interplay of recombination and selection. Evolution. 2022;76(4):782–98.

Publisher's Note