Revised Proof

Journal of Genetics _####################_
https://doi.org/10.1007/s12041-022-01398-3

© Indian Academy of Sciences

RESEARCH ARTICLE

Remnants of *SIRE1* retrotransposons in human genome?

BUKET CAKMAK GUNER[1,2]*, ELIF KARLIK[1,3]* and NERMIN GOZUKIRMIZI[1]

[1]*Department of Molecular Biology and Genetics, Institute of Science, Istanbul University, 34134 Vezneciler, Istanbul, Turkey*
[2]*Department of Molecular Biology and Genetics, Faculty of Art and Science, Avrasya University, 61000 Trabzon, Turkey*
[3]*Department of Molecular Biology and Genetics, Faculty of Engineering and Natural Sciences, Istinye University, 34010 Istanbul, Turkey*
*For correspondence. E-mail: Buket Cakmak Guner, buket.cakmak@avrasya.edu.tr; Elif Karlik, elifkarlik@istinye.edu.tr.

Received 14 January 2022; revised 10 August 2022; accepted 22 August 2022

Abstract. Evolution is unaimed changes in time that a genome is shaped by a collection of random mutations, recombination, integrations, and reorganizations. Transposable elements (TEs) are mobile fragments representing a major portion of most eukaryotic genomes, and are therefore considered as a key player in evolution. They are one of the main sources of genetic variability and have a large impact on genome structure and stability in eukaryotes. In this study, the plant *SIRE1* retrotransposon insertions were demonstrated in the human genome by using barley *SIRE1* interretrotransposon amplified polymorphism PCR (IRAP-PCR) primers. According to the IRAP-PCR analysis, different distribution patterns were observed for 24 participants used in this study. The polymorphism ratios of *SIRE1* were calculated, and among all samples they were detected between 0 to 38%. Similarly, internal domains and LTR sequences of *SIRE1* were investigated by sequencing. Partial *GAG, RT* and *ENV* gene sequences were detected in the human genome by performing sequence and bioinformatic analyses. According to the bioinformatic analysis, partial *SIRE1* ENV sequences were interestingly detected in both human and chimpanzee chromosome 1. Partial *SIRE1* ENV sequences in chromosome 1 were also found to be associated with neuroblastoma breakpoint family members' (*NBPFs*) in humans. Polymorphic TE insertions in the human genome may be an essential source of natural genetic variation with subtle effects on genome regulation, providing considerable source material for ongoing human evolution.

Keywords. evolution; transposable elements; retrotransposons; human genome; *SIRE1*.

Introduction

Evolution of life is unaimed hereditary changes relayed through time (Dawkins 1974; Laland *et al.* 2015). The genome is formed by a collection of random mutations, recombination, integrations, and reorganizations that are shaped by natural selection (Skern-Mauritzen and Mikkelsen 2021). Additionally, the success of species strictly relies on the composition and functionality of their genome. Only a small fraction of nuclear DNA represents a coding that on the contrary the genome is mostly composed of repetitive DNA and in particular of transposable elements (TEs). TEs, or transposons, are found in almost all plants, animals and even humans that can jump from site to site in the genome during the life cycle of a cell. Some transposons encode the enzymes which perform their excision (Lander *et al.* 2001; de Koning *et al.* 2011). TEs have different shapes, lengths, and different ways of mobilization. In the genomes, TEs are normally nonrandomly distributed and considered as the main drivers of genome evolution. Therefore, they have diverse impacts on genome evolution such as promoting genome size expansion influencing genome plasticity and instability, also can generate new coding genes or regulatory elements modulating gene and chromosomal rearrangements and contribute to horizontal gene transfers (summarized by Biscotti *et al.* 2019; Viviani *et al.* 2021).

TEs are generally classified into two main classes; class I: retrotransposons use a 'copy and paste' mechanism for their replication and expansion. Class II: DNA transposons use a 'cut and paste' mechanism (Rebollo *et al.* 2012).

Buket Cakmak Guner and Elif Karlik contributed equally to this work.

Revised Proof

Interestingly, class I elements are mostly found in eukaryotic lineages, but rarely in prokaryotes. However, class II elements are found in both prokaryotes and eukaryotes, suggesting that progenitors of both classes were presumably present in the common ancestor of all eukaryotes (Sotero-Caio *et al.* 2017). Retroviruses are estimated to be inserted into the mammalian lineages about 550 million years ago (Hayward 2017). Older endogenous virus sequences may still exist in human genome, but they cannot be recognized according to the accumulation of mutations, including deletions, insertions, or homologous recombination events with large deletions (Moelling and Broecker 2019). Retrotransposons are also subdivided into two categories; long terminal repeat (LTR)-retrotransposons including human endogenous retrovirus (HERVs) (Mills *et al.* 2007; Blomberg *et al.* 2009), and non-LTR retrotransposons including long-interspersed elements (LINEs), short-interspersed elements (SINEs) and in humans, SINE-VNTR-Alu elements (SVAs). Two-thirds of the human genome is composed of more or less intact ERVs and related retrotransposons (de Koning *et al.* 2011). LINEs are considered as the most ancient elements known as autonomous retrotransposons while SINEs and SVA elements are nonautonomous and depend on LINE-1 machinery for retrotransposition (Zimmerly and Semper 2015). Additionally, LINE is considered to be responsible for most reverse transcription events, including the retrotransposition of the nonautonomous SINEs and the emergence of processed pseudogenes (Esnault *et al.* 2000; Wei *et al.* 2001).

LTR-retrotransposons are the autonomous elements containing the *gag* and *pol* genes that encode a protease, reverse transcriptase, RNAse H and integrase. Interestingly, retrotransposons have distinct evolutionary histories. Some researchers considered that exogenous retroviruses have arisen from endogenous retrotransposons by the acquisition of a cellular envelope gene *(env)* (Malik *et al.* 2000). Some others claimed that LTR endogenous retroviruses seem to evolve from ancient viral infections of the germline and are maintained vertically. Endogenous retroviruses (ERVs) encode Pro, Gag, Pol, and sometime Env-like proteins like their exogenous cousins (Lander *et al.* 2001). Mammalian retroviruses are divided into three classes (I±III), containing many families with independent origins. Most of the LTR retrotransposons (85%) are 'fossils' consisting only of LTRs (Mager and Stoye 2015). TE activity, especially ERVs, has reduced in the hominid lineage relative to other mammalian lineages (Lander *et al.* 2001). Moreover, Alu, LINE-1, SVA retrotransposons and ERVs, including human ERVs are active mobile genetic elements in the human genome (Kazazian *et al.* 1988; Batzer and Deininger 1991; Brouha *et al.* 2003; Ostertag *et al.* 2003; Wildschutte *et al.* 2016; Karlik *et al.* 2021). Although they were assumed as junk DNA, later it was understood that these elements both could shape the genomes and play substantial roles in the regulation of human genes (Feschotte 2008; Rebollo *et al.* 2012; Chuong *et al.* 2017), and were also associated with diseases (Wang *et al.* 2017; Karlik *et al.* 2021; Mao *et al.* 2021). Therefore, understanding the functions of TEs are required. Most of the studies on mammalian TEs have been focussed on human, mice, and chimpanzee ERV and LINE families of TEs (Yohn *et al.* 2005; Marchetto *et al.* 2013; Mun *et al.* 2014).

Throughout the evolutionary history of primate TEs, some TEs are conserved over multiple phylogenetic clades and orders while others are restricted to particular lineages (Ward *et al.* 2018). The transmission of DNA between organisms is demonstrated to be not necessarily closely related to horizontal transfer (Soucy *et al.* 2015). One type of horizontal transfer among eukaryotes is widespread though, that of TEs, however, the frequency, impact, and mechanisms underlying these transfers have remained more obscure in eukaryotes (Martin 2017). Recently, studies indicated that chromosomal distribution of TEs in natural plant populations were found to be closely related species, suggesting that a processor of some TE clusters may be ancient and shared between species (Wright *et al.* 2001; Altinkut *et al.* 2006). One of the first cases reported in eukaryotes for horizontal transfer of TEs was the P elements (Daniels *et al.* 1990). The study demonstrated that P elements were distributed within a subgenus of *Drosophila*, and are almost identical in *D. melanogaster* and *D. willistoni* despite the >26 million years of divergence separating the two species, indicating that TEs can cross species barriers has deeply transformed our understanding of both TE evolutionary dynamics and host genome evolution (Daniels *et al.* 1990; Peccoud *et al.* 2018).

Another retrotransposon is Sireviruses, an ancient retrotransposon among *Tyl/copia* elements in plants, and has a unique genome structure (Gao *et al.* 2003; Bousios *et al.* 2010, 2012). Sequence studies of Sireviruses demonstrated that highly conserved sequence motifs have been found within the extremely divergent noncoding domains known to participate in the life cycle of LTR retrotransposons (Bousios *et al.* 2010). Sireviruses are named as the *SIRE1* element after it was identified in soybean. Two notable features of Sireviruses differentiate it from the other *Ty1/copia* genera is that they consist of an *env*-like gene after *pol* and, encode a significantly larger *Gag* protein (Peterson-Burch and Voytas 2002; Havecker *et al.* 2005). Each copy of *SIRE1* is ∼11 kb, making *SIRE1* one of the largest retrotransposons in soybean, additionally, *SIRE1* is present and active in other plant species, including barley, rice, maize, wheat, legumes, beets, bananas, and agaves (Laten and Morris 1993; McCarthy *et al.* 2002; Holligan *et al.* 2006; Weber *et al.* 2010; Hřibova *et al.* 2010; Bousios *et al.* 2012; Cakmak *et al.* 2015).

To determine the activity of a retrotransposon, there are many retrotransposons-based molecular marker techniques developed (Gozukirmizi *et al.* 2015; 2016), thus one of the most used techniques is interretrotransposon amplified polymorphism (IRAP) molecular marker technique. In this technique, two flanked retrotransposons or solo LTR are amplified by special IRAP primers. According to the absence or presence of the PCR product and the absence of

Revised Proof

amplification indicates lack of the retrotransposon at a specific region, the polymorphism rates are calculated, commonly used to study the genetic diversity in plants and even humans (Kalendar *et al.* 1999; Schulman and Kalendar 2005; Guliyev *et al.* 2013; Yuzbasioglu *et al.* 2016; Cakmak *et al.* 2017; Cakmak-Guner *et al.* 2018; Karlik *et al.* 2021). In recent years, there are many reports to demonstrate the analysis of *SIRE1* element, especially in plants (Martín Sanz *et al.* 2007; Du *et al.* 2010; Bousios *et al.* 2012; Cakmak *et al.* 2015). Recently, researchers demonstrated that *SIRE1* and *BARE-1* are active retroelements in horse and sheep genomes and they have shown polymorphism among individuals (Elkina *et al.* 2015). Additionally, our research group also investigated the transferability of the barley retrotransposon *Sukkula* primers for human genome (Cakmak *et al.* 2017). Thus, our research group detected *Sukkula* insertion polymorphisms of the participants used in the study. Therefore, the aim of this study was the determine the *SIRE1* elements in human genome. Polymorphism rates among human genome belonging to different ages were analysed by IRAP method. Additionally, we performed sequence analysis of *LTR* sequences and internal domains (*GAG, ENV* and *RT*) of *SIRE1* element. Our findings indicated that remnants of *SIRE1* may possibly be in the human genome.

Materials and methods

Samples

In this study, we obtained genomic DNA samples from Dr Kaniye Sahin from the Department of Molecular Biology and Genetics, Istanbul University. Genomic DNAs were extracted with DNA extraction kit (Roche Diagnostics GmbH, Mannheim, Germany). This study was carried out according to the ethical standards. The quality of the samples were detected with Spectrophotometric and electrophoretic analyses. A total of 24 DNA samples (12 females and 12 males) within the age range of 10–79 years (table 1) were used in the study. Individuals who participated in this study were not related to each other and did not have any diseases.

Table 1. Participant's information.

Sample numbers*	Age range (years)	Sample numbers*	Age range (years)
1	10–19	13	40–49
2	10–19	14	40–49
3	10–19	15	40–49
4	10–19	16	40–49
5	10–19	17	60–69
6	20–29	18	60–69
7	20–29	19	60–69
8	20–29	20	60–69
9	20–29	21	70–79
10	20–29	22	70–79
11	20–29	23	70–79
12	20–29	24	70–79

*Total 24 (12 females and 12 males) samples within the age range of 10–79 years numbered randomly in order.

IRAP analysis

The primer sequence used for IRAP was 5′-CAGTTATG-CAAGTGGGATCAGCA-3′ as obtained from Chesnay *et al.* (2007). PCR reaction was carried out by using 2× Sapphire enzyme mix (Takara, RR350A). Concentration and quantity of PCR components were modified from our previous studies associated with this work (Cakmak *et al.* 2017; Cakmak-Guner *et al.* 2018). In addition, PCR conditions were also modified from the same previous works. The reaction was carried out with 30 cycles. Initial denaturation and denaturation temperature were 94°C; annealing temperature was 51°C and extension and final extension temperature were 72°C. PCR products run 2% agarose gel electrophoresis in 1× TBE buffer at 150 V for 120 min. Band sizes of amplicons were determined with GeneRuler^{TM} 100-bp plus marker (SM0321, Fermentas) and gel was photographed on a UV transilluminator.

SIRE1-LTR analysis

In addition to IRAP analysis, primers of *SIRE1*-LTR sequences were obtained from Laten *et al.* (2003). Forward primer was 5′-TGGAAGGTTGTAAACAGTGGC-3′ and reverse primer was 5′-ATATTTTCGGGCAGATG-3′. Total PCR reaction volume was 20 μL and final concentration of PCR components were as follow: 1× buffer, 2.5 mmol/L $MgCl_2$, 0.2 mmol/L dNTP mixture, 0.5 μM/μL of each primer, 0.05 U/μL Dream Taq PCR Enzyme Mix (EP0702, Thermo Scientific) and 4 ng/μL genomic DNA. PCR reaction was performed with 30 cycles. Initial denaturation (30 s) and denaturation temperature were 94°C (3 min); annealing temperature was 52°C (30 s) and extension (2 min) and final extension temperature were 72°C (10 min). PCR products were mixed 6× loading buffer and run in 2% agarose gel electrophoresis for 100 min in 1× TBE buffer. Band sizes of amplicons were determined with GeneRuler^{TM} 100 bp plus marker (SM0321, Fermentas) and the gel was photographed on a UV transilluminator.

Polymorphism determination

The polymorphism rates were determined based on the Jaccard similarity coefficient (Jaccard 1908). Moreover, band profiles were also analysed by GelJ v.2.0 to construct the phylogenetic tree. In addition to Jaccard's coefficient,

Revised Proof

samples were clustered based on the band distances with the unweighted pair-group method with arithmetic mean (UPGMA) clustering method to measure of dissimilarity, thus avoiding characterizing the dissimilarity by extreme values (minimum and maximum) between the considered genotypes (Cruz *et al.* 2014; Heras *et al.* 2015).

Internal domain analysis

SIRE1 GAG, ENV, and RT internal domains were also investigated in DNA samples using suitable primers (Laten *et al.* 2003) (in table 2, primer 1–2 for ENV, 3–4 for GAG and 5–6 for RT domain). PCR components and PCR conditions were the same as *SIRE1*-LTR analysis. Annealing at different temperatures for GAG, ENV, and RT analyses are indicated in table 2.

Sequencing and bioinformatic analysis

The Sanger sequencing was performed using the primers used for internal domain analysis indicated in table 2. Afterwards, the retrieved sequences were used as a query and subjected to homology search with the available human genome at Ensembl database, additionally, the retrieved sequences were also used as a query and subjected to homology search with the available primate's genomes at Ensembl database using BLASTN tool (Altschul *et al.* 1990). Nucleotide multiple sequence alignment was performed using the Muscle algorithm (Edgar 2004). MEGA 7 was used to calculate phylogeny reconstruction based on the maximum-likelihood (ML) model (Felsenstein 1981), including the following parameters: nucleotide, Tamura–Nei model (Tamura and Nei 1993); data subset to use, complete deletion; bootstrap analysis with 1.000 replicates. Tree inference options: nearest-neighbour-interchange (NNI); codon usage: 1st+2nd+3rd+noncoding (Kumar *et al.* 2016). The procedures followed were in accordance with the current ethical standards.

Table 2. Primer sequences used in this study.

	Name	Sequence (5′–3′)	(°C)
1	*SIRE-ENV*/F	ACATTGTCTCGACACAGGG	
2	*SIRE*-3′LTR/R	ATATTTTCGGGCAGATG	52
3	*SIRE*-5′LTR/F	TGGAAGGTTGTAAACAGTGGC	
4	*SIRE-GAG*/R	AGTCGAAAGGGATGTTCCG	52
5	*SIRE-RT*/F	GAGGCACTGACTGATGAGTTC	
6	*SIRE-RT*/R	TTCTTTGCATACTTGCTTTGTGAG	47

Results and discussions

Display of SIRE1 polymorphisms in participants

TEs are mobile DNA fragments that can move from one place to another on their host chromosomes or genome or pass to another genome of the organism which is referred as the horizontal transfer of the genetic material. Especially, retrotransposons were found in almost every species, playing an important role in both prokaryotic and eukaryotic evolutions (Schaack *et al.* 2010). In this study, we aimed to find the possible remnants of plant LTR retrotransposon, *SIRE1* element, in the human genome. The study demonstrated that *SIRE1* primers can also be transferred to the human TE studies according to amplification of internal domains and LTR sequences of *SIRE1* (see figures 1–6). To evaluate *SIRE1* presence, we performed IRAP-PCR analysis, the results showed that *SIRE1* primers were worked out in the human genome (see in figure 1). We observed two main clusters as a result of GelJ analysis (figure 2). The first group consisted of six subjects (no: 1, 11, 12, 22, 23, 24), the other 17 subjects were found in the second group. We also evaluated monomorphic and polymorphic bands among samples to calculate polymorphism percentages. Three hundred and twenty-three monomorphic and 61 polymorphic scorable bands were determined ranging from 400 to 2000 bp. Absent (−) or present (+) bands are shown in table 3. There were 0–38% polymorphism ratios among all samples. We also obtained % 0–38 polymorphism among females (12 females) and % 0–31 in males (12 males). In addition, differences between males and females were also found to be 0–38% (table 4). We had five different groups in terms of the average range, including 24 subjects. Two subjects of each group were selected for *SIRE1* LTR analysis (subject no: 1, 2, 11, 12, 14, 15, 17, 18, 21 and 23). Band profiles of IRAP-PCR among samples were the same, ranging from 250 to 2000 bp (figure 3), while the *SIRE1* LTR analysis results were observed between 500 and 2000 bp. Based on the LTR band profiles analysis, two subjects of each group were randomly selected for internal domain analysis (subject no: 1, 2, 11, 12, 14, 15, 17, 18, 21 and 23). Band profiles among samples are shown in figures 4, 5, 6. Also, we observed different sequences sizes for all internal domains. We observed RT sequences nearly between 300 and 2000 bp; and GAG sequences nearly between 150 and 2000 bp and ENV sequences nearly between 500 and 3000 bp.

It has been a long time that the origin of retrotransposons is being discussed, that retrotransposons are the remnants of ancient retroviral germline infections. In time, these insertions are considered as evolutionarily fixed in the genome. Approximately 450.000 HERVs consists of 8% of the human genome, containing hallmark retroviral elements like the *gag*, *pol*, *env* genes and LTRs that act as promoters (Lander *et al.* 2001). This structure is similar to the retrotransposons that *SIRE1* also comprises *gag*, *pol*, *env*-like

Revised Proof

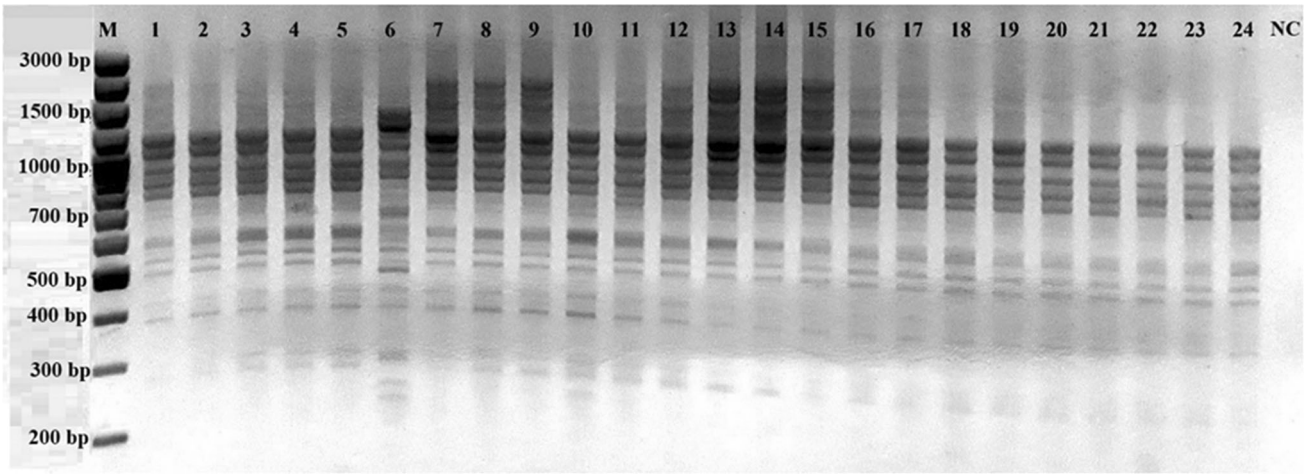

Figure 1. *SIRE1* IRAP-PCR results of participants. NC, negative control; M, 1-kb marker and lane numbers represent the total 24 (12 females and 12 males) samples numbered in order according to table 1.

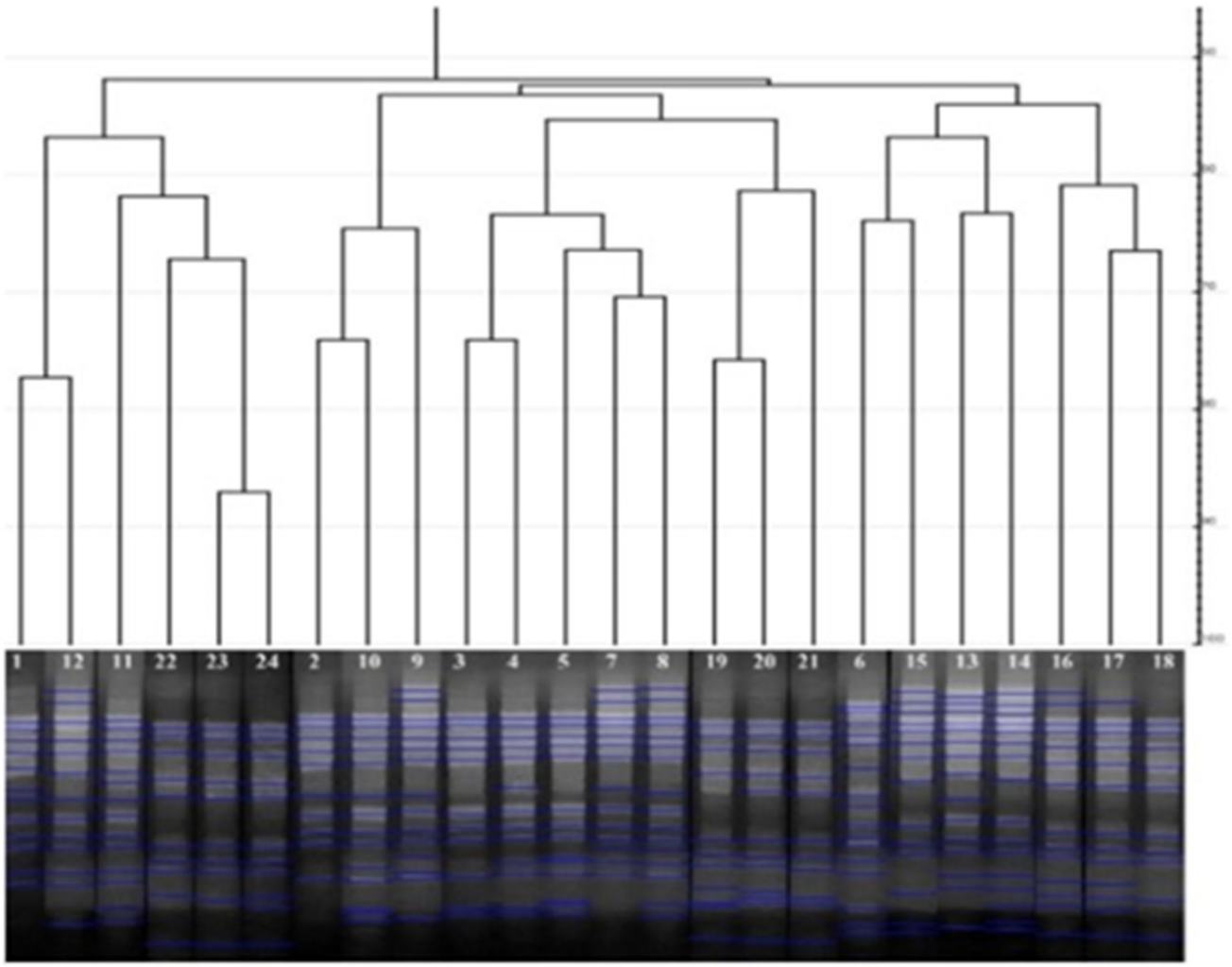

Figure 2. *SIRE1* UPGMA results (GelJ v.2.0 analysis). Lane numbers represent the total 24 (12 females and 12 males) samples numbered in order according to table 1.

Revised Proof

Table 3. Absent (−) and present (+) bands of *SIRE1*.

Samples*	+	−	Samples*	+	−
1	12	4	13	16	0
2	12	4	14	16	0
3	12	4	15	16	0
4	10	6	16	14	2
5	12	4	17	14	2
6	14	2	18	12	4
7	15	1	19	12	4
8	16	0	20	12	4
9	16	0	21	12	4
10	14	2	22	12	4
11	14	2	23	12	4
12	16	0	24	12	4

+, Monomorphic; −, polymorphic bands number.
*Total 24 (12 females and 12 males) samples numbered in order.

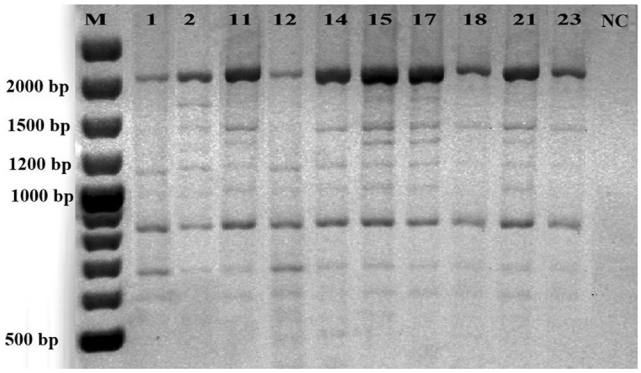

Figure 3. *SIRE1* LTR results. Lane numbers represent the total 24 (12 females and 12 males) samples numbered in order according to table 1.

genes (Cakmak *et al.* 2015). Additionally, studies exhibited that the same retrotransposons are mostly found in the different species and probably occurred as a result of horizontal gene transfer mechanism. The transfer of new genes from different species is especially important for prokaryotic evolution. A few cases have also been reported that horizontal gene transfer mechanism play substantial roles in eukaryotic evolution (reviewed in Gozukirmizi *et al.* 2015). *SIRE1* may have integrated into the human genome as a retrovirus in time due to the similarities between ERVs and *SIRE1*. However, identifying horizontal gene transfer events faces some difficulties, including the variable quality of available genome sequences and limitations of analytical procedures. One of the main types of horizontal transfer cases is the amplification of TEs by transposition in the receiving genome (Peccoud *et al.* 2018). For the last two decades, the cases that crossed species boundaries by both

Table 4. *SIRE1* polymorphism percentages (%).

Sample numbers*																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	–																							
2	0	–																						
3	0	0	–																					
4	17	17	17	–																				
5	0	0	0	17	–																			
6	14	14	14	29	14	–																		
7	20	20	20	22	20	7	–																	
8	25	25	25	38**	25	13	6	–																
9	25	25	25	38**	25	13	6	0	–															
10	14	14	14	29	14	0	7	13	13	–														
11	14	14	14	29	14	0	7	13	13	0	–													
12	25	25	25	38**	25	13	6	0	0	13	13	–												
13	25	25	25	38**	25	13	6	0	0	13	13	0	–											
14	25	25	25	38**	25	13	6	0	0	13	13	0	0	–										
15	25	25	25	38**	25	13	6	0	0	13	13	0	0	0	–									
16	14	14	14	29	14	0	7	13	13	0	0	13	13	13	13	–								
17	14	14	14	29	14	0	7	13	13	0	0	13	13	13	13	0	–							
18	0	0	0	17	0	14	20	25	25	14	14	25	25	25	25	14	14	–						
19	0	0	0	17	0	14	20	25	25	14	14	25	25	25	25	14	14	0	–					
20	8	8	8	25	8	21	27	31	31	21	21	31	31	31	31	21	21	8	8	–				
21	8	8	8	25	8	21	27	31	31	21	21	31	31	31	31	21	21	8	8	0	–			
22	8	8	8	25	8	21	27	31	31	21	21	31	31	31	31	21	21	8	8	0	0	–		
23	8	8	8	25	8	21	27	31	31	21	21	31	31	31	31	21	21	8	8	0	0	0	–	
24	8	8	8	25	8	21	27	31	31	21	21	31	31	31	31	21	21	8	8	0	0	0	0	–

*Total 24 (12 females and 12 males) samples numbered in order.
**Percentages (%) have been calculated using Jaccard similarity coefficient. Numbers marked with dark indicated highest percentages (%).

Revised Proof

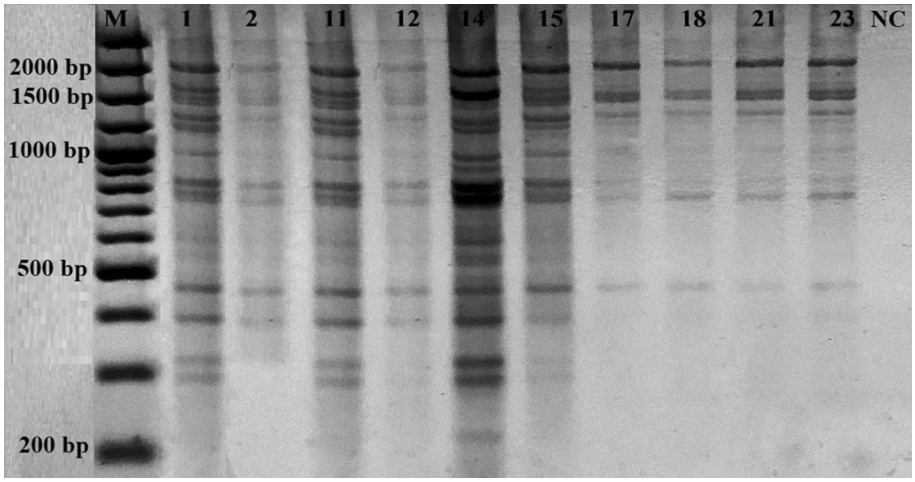

Figure 4. RT domain results. Lane numbers represent the total 24 (12 females and 12 males) samples numbered in order according to table 1.

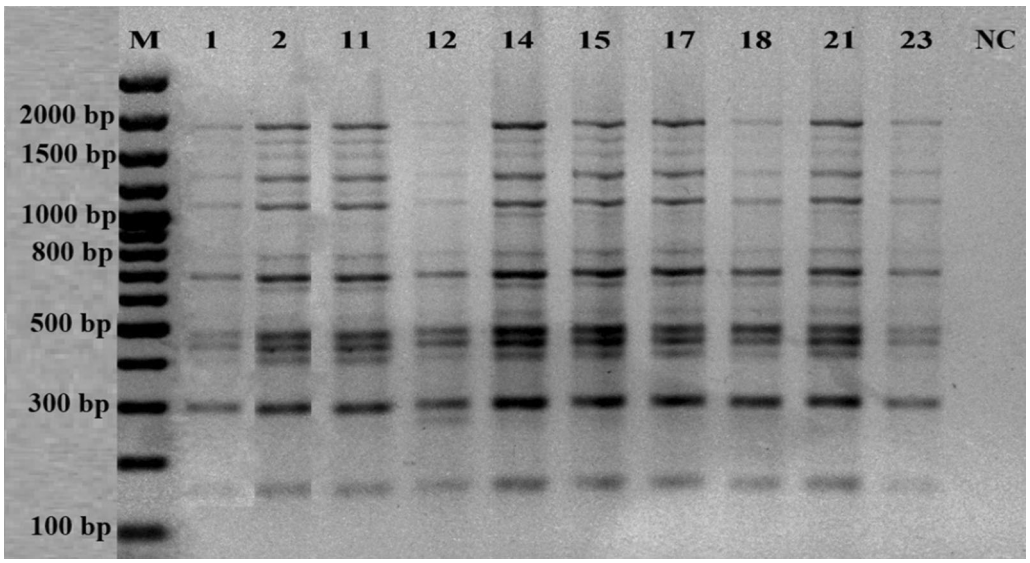

Figure 5. Gag domain results. Lane numbers represent the total 24 (12 females and 12 malse) samples numbered in order according to table 1.

RNA-mediated and DNA-mediated elements have been reported on many occasions, totally 218 convincing cases of horizontal transfer of TEs with 103, 97, and 14 cases affecting DNA transposons, LTR retrotransposons, and non-LTR retrotransposons, respectively (Schaack *et al.* 2010). However, studies of plant-based transposons in animal genomes are rare. To date, there was one known study that investigates plant-specific *BARE-1* and *SIRE1* retrotransposons in the genome of farm animals (Elkina *et al.* 2015). Cakmak *et al.* (2017) identified another plant-specific retrotransposon *Sukkula* in the human genome. They found the polymorphism rates as 8–100% among all samples. Most researchers stated that horizontal transfer not only clarifies the continuousness of the TEs over the evolutionary times despite host defense mechanisms against TEs, but it may also explain that horizontal transfer may contribute to the composition and evolution of eukaryote genomes (Schaack *et al.* 2010; Wallau *et al.* 2012; Ivancevic *et al.* 2013).

Evolutionary analysis of SIRE1 and other different species

Over half of primate genomes are composed of TEs (Jurka 2000; de Koning *et al.* 2011). Comparison sequence similarity between the human and chimpanzee genomes demonstrated that most of the TEs are present in both species (Chimpanzee Sequencing and Analysis Consortium 2005; Ramsay *et al.* 2017). There was strong evidence in the recent studies indicating chromosomal distributions are an ongoing intrapopulation activity of both hAT and CACTA TE families (Wright *et al.* 2001; Altinkut *et al.* 2006). Thus, we intended to find out the intraspecific and interspecific

Revised Proof

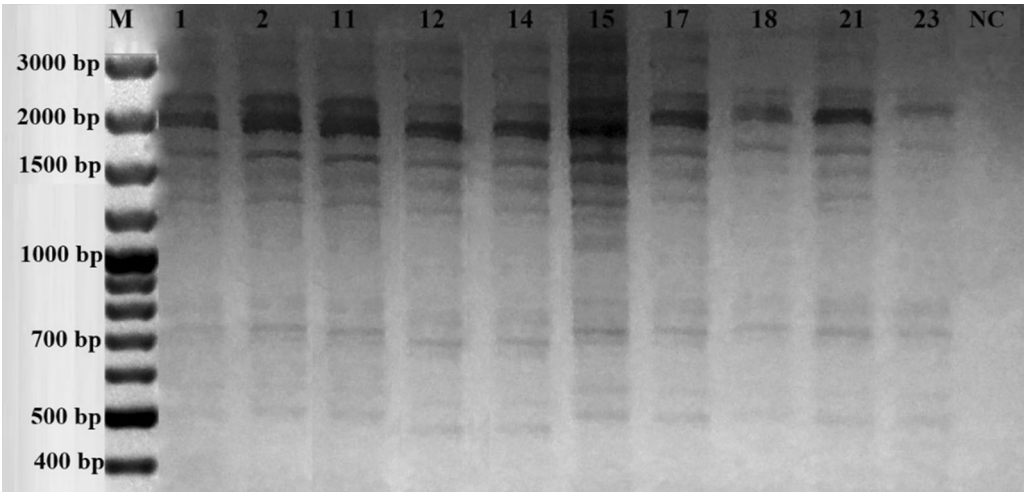

Figure 6. Env domain results. Lane numbers represent t the total 24 (12 females and 12 males) samples numbered in order according to table 1.

similarities/differences in patterns of *SIRE1* TE that we have previously identified the soybean *SIRE1* sequences in barley, and concluded that *SIRE1* was active (Cakmak *et al.* 2015). In this study, we used *SIRE1* TEs primers, which were designed according to barley *SIRE1* sequence analysis (Cakmak *et al.* 2015), to demonstrate the *SIRE1* polymorphism among human genome, thus we were able to analyse and calculate the *SIRE1* polymorphism, additionally, to be able to re-sequence the partial *SIRE1* sequences from human genome. The sequence analysis demonstrated that the partial *SIRE1* LTR and GAG sequences were found to be located in human chromosomes 4, 9, 20 and 22 for LTRs and chromosome 8 for GAG.

TEs adapted different strategies to ensure their evolutionary survival. While LINEs and SINEs rely almost exclusively upon vertical transmission within the host genome, DNA transposons depend on relatively frequent horizontal transfer (Malik *et al.* 2000). LTR retrotransposons use both strategies, with some being long-term active residents of the human genome (such as members of the ERV family) and others having only short residence times (Lander *et al.* 2001). *SIRE1* may have integrated into the human genome over 550 million years as a retrovirus in time. Some researchers point out that infection and integration are unique events occurring at a fast pace (Wolf and Koonin 2013; Moelling and Broecker 2019). However, accumulation of point mutations and homologous recombination events may lead to gene loss or gene reductions. Recent studies demonstrated that repetitive sequences in the human genome contributed to human genome evolution by becoming functional elements, such as protein coding regions and binding sites for transcriptional regulators (Sorek *et al.* 2002; Bejerano *et al.* 2006; Kojima 2018). These repeats are shared with the genomes of many other mammals, vertebrates, and amniotes that nearly all TE families are shared between chimpanzees and humans with an exception, which is the ERV family *PtERV1*, is present in the genomes of chimpanzees and gorillas but not in humans (Yohn *et al.* 2005; Kojima 2018). However, the evolutionary history of TEs in a variety of eukaryote lineages the ERVs of vertebrates can be readily affiliated with TEs and, like them, are capable of spreading vertically and horizontally (Belshaw *et al.* 2004; Zhuo and Feschotte 2015). In this study, we also found that partial *SIRE1* sequences are shared between human and chimpanzee according to bioinformatic analysis. Additionally, we demonstrated that these partial *SIRE1* ENV sequences are shared with the genomes of other primates. However, the partial *SIRE1* ENV sequences are mostly concentrated in human chromosome 1 which is the largest human chromosome (Murphy *et al.* 2003). Additionally, partial *SIRE1* ENVs were also shown to be placed in human chromosomes 2, 4, 6, 7, 9, 11, 13 and sex chromosome X. According to chimpanzee BLAST results, partial *SIRE1* GAG also located in chimpanzee chromosome 8, although partial *SIRE1* LTRs located in chimpanzee chromosomes 4, 7, 9 and 20. Similarly, partial *SIRE1* ENVs were found to be mostly concentrated in chimpanzee chromosome 1. However as distinct from human, partial *SIRE1* ENVs were located in chimpanzee chromosomes 11, 12, 13, 22 and sex chromosome X. Both human and chimpanzee partial *SIRE1* ENVs sequences in chromosome 1 were found to be associated with neuroblastoma breakpoint family members' (*NBPF*s) etc. *NBPF3* for chimpanzee and *NBPF8*, *NBPF10*, *NBPF14*, *NBPF19*, *NBPF20* and *NBPF26* for human (figure 7). Interestingly, the shared chromosomes were different between human and chimpanzee for the partial *SIRE1* LTR and ENV sequences in this study. Nearly all studies have focussed on human TE-derived sequences associated with regulatory elements. Thus, studies indicated they are remnants of proportionally ancient insertion events and no longer capable of transposition (Wang and Jordan 2018). In our study, we also observed *SIRE1* retrotransposon

Revised Proof

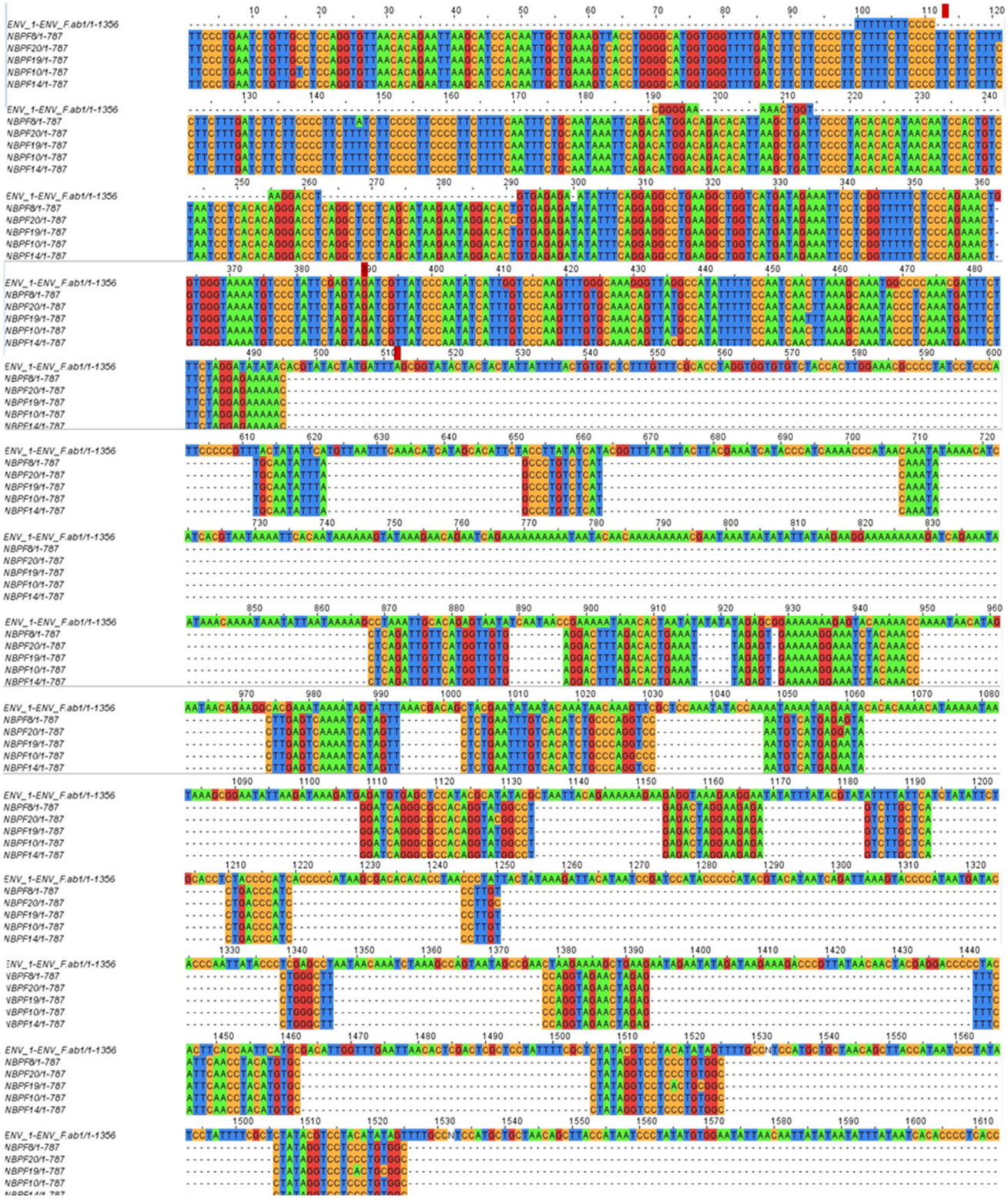

Figure 7. Display of sequence alignment of human *NBPF* genes and the sequenced human *SIRE1* sequence.

polymorphism events in different individuals. After fixation of *SIRE1* into human genome, point mutations, duplication, and homologous recombination events led to this polymorphism in time.

Gene families essentially emerge by two main gene duplication mechanisms, unequal crossing-over and retrotransposition (Lynch 2007). Tandem repeats are created by the first mechanism. However, second mechanism is the

Revised Proof

insertion of an intronless cDNA (Eirín-López *et al.* 2012) which is highly similar to retrotransposition mechanism used by retrotransposons. *NBPF* is a gene family with a complicated genomic organization, expanding during recent primate evolution. Moreover, they are brain associated genes, suggesting at least one of them may supress the development of neuroblastoma and probably of other tumour types (Vandepoele *et al.* 2005, 2008, 2009; Gregory *et al.* 2006; Popesco *et al.* 2006; Vandepoele and van Roy 2007; Diskin *et al.* 2009). Previous bioinformatics studies have reported *NBPF* have tandem repeats in these genes, which is also called *NBPF* repeats, based on diverging; 1.5–1.6-kb repeat units (Altschul *et al.* 1990; Benson 1999; Vandepoele *et al.* 2005; Gregory *et al.* 2006; Gelfand *et al.* 2007; Warburton *et al.* 2008). The copy number of these genes in human is determined to be variable (Tuzun *et al.* 2005; Redon *et al.* 2006). However, a remarkable reduction of *NBPF* copy number was demonstrated in other primates, additionally, the *NBPF* copies are absent in mouse (Vandepoele *et al.* 2005, 2009; Popesco *et al.* 2006). According to our sequencing and bioinformatic analysis, both human and chimpanzee partial *SIRE1* ENVs sequences in chromosome 1 were found to be related to *NBPF* members' etc. *NBPF3* for chimpanzee and *NBPF8*, *NBPF10*, *NBPF14*, *NBPF19*, *NBPF20* and *NBPF26* for human (figure 7), indicating *SIRE1* ENVs sequences may be associated with these *NBPF* copies and recent primate evolution.

Conclusion

Viruses may contribute to the evolution of life since 550 million years ago (Hayward 2017). Viruses can be admitted as drivers of evolution today, while they are mostly considered as pathogens (Villarreal and Witzany 2010; Moelling and Broecker 2019). However, origin of retrotransposons is estimated that retrotransposons are the remnants of ancient retroviruses (Lander *et al.* 2001). Understanding their long-term impacts of integration and fixation of TEs is substantial to the evolution of eukaryotic genomes. Based on the polymorphism and bioinformatics studies, our study was only able to demonstrate the plant *SIRE1* retroelement sequences—*SIRE1* LTR, RT, GAG and ENV domains—in the human genome. However, polymorphic human TE insertion variants contain a substantial source of natural genetic variation with subtle effects on genome regulation and human health, providing considerable source material for ongoing human evolution.

Acknowledgement

Scientific Research Projects Coordination Unit of Istanbul University (grant number FDK-2017-22142) has supported this work.

References

Altinkut A., Kotseruba V., Kirzhner V. M., Nevo E., Raskina O. and Belyayev A. 2006 Ac-like transposons in populations of wild diploid *Triticeae* species: comparative analysis of chromosomal distribution. *Chromosome Res.* **14**, 307–317.

Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Batzer M. A. and Deininger P. L. 1991 A human-specific subfamily of Alu sequences. *Genomics.* **9**, 481–487.

Bejerano G., Lowe C. B., Ahituv N., King B., Siepel A., Salama S. R. *et al*. 2006 A distal enhancer and an ultraconserved exon are derived from a novel retrotransposons. *Nature.* **441**, 87–90.

Belshaw R., Pereira V., Katzourakis A., Talbot G., Paces J., Burt A. *et al*. 2004 Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. USA* **101**, 4894–4899.

Benson G. 1999 Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580.

Biscotti M. A., Carducci F., Olmo E. and Canapa A. 2019 Vertebrate Genome Size and the Impact of Transposable Elements in Genome Evolution. In *Evolution, Origin of Life Concepts and Methods* (ed. P. Pontarotti), Springer, Cham, pp. 233–251.

Blomberg J., Benachenhou F., Blikstad V., Sperber G. and Mayer J. 2009 Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. *Gene* **448**, 115–123.

Brouha B., Schustak J., Badge R. M., Lutz-Prigge S., Farley A. H., Moran J. V. *et al*. 2003 Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA* **100**, 5280–5285.

Bousios A., Darzentas N., Tsaftaris A. and Pearce S. R. 2010 Highly conserved motifs in non-coding regions of Sirevirus retrotransposons: the key for their pattern of distribution within and across plants. *BMC Genomics* **11**, 89.

Bousios A., Minga E., Kalitsou N., Pantermali M., Tsaballa A. and Darzentas N. 2012 MASiVEdb: the sirevirus plant retrotransposon database. *BMC Genomics* **13**, 158.

Cakmak B., Marakli S. and Gozukirmizi N. 2015 *SIRE1* Retrotransposons in barley (*Hordeum vulgare* L.). *Russ. J. Genet.* **51**, 661–672.

Cakmak B., Marakli S. and Gozukirmizi N. 2017 *Sukkula* retrotransposon movements in the human genome. *Biotechnol. Biotechnol. Equip.* **31**, 900–905.

Cakmak-Guner B., Karlik E., Marakli S. and Gozukirmizi N. 2018 Detection of HERV-K6 and HERV-K11 movements in human genome. *Biomed Rep.* **9**, 1.

Chesnay C., Kumar A. and Pearce S. R. 2007 Genetic diversity of *SIRE1* retroelements in annual and parennial glycine species revealed using SSAP. *Cell. Mol. Biol. Lett.* **12**, 103–110.

Chimpanzee Sequencing and Analysis Consortium 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87.

Chuong E. B., Elde N. C. and Feschotte C. 2017 Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86.

Cruz C. D., Salgado C. C. and Bhering L. L. 2014 Biometrics applied to molecular analysis in genetic diversity. *Plant Breed. Biotechnol.* 47–81.

Daniels S. B., Peterson K. R., Strausbaugh L. D., Kidwell M. G. and Chovnick A. 1990 Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **124**, 339–355.

Revised Proof

Dawkins R. 1974 *The Selfish Gene*. Oxford University Press, New York.

de Koning A. P., Gu W., Castoe T. A., Batzer M. A. and Pollock D. D. 2011 Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384.

Diskin S. J., Hou C., Glessner J. T., Attiyeh E. F., Laudenslager M., Bosse K. *et al*. 2009 Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459**, 987–991.

Du J. C., Grant D., Tian Z. X., Nelson R. T., Zhu L. C., Shoemaker R. C. *et al*. 2010 SoyTEdb: a comprehensive database of transposable elements in the soybean genome. *BMC Genomics* **11**, 113.

Edgar R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Eirín-López J. M., Rebordinos L., Rooney A. P. and Rozas J. 2012 The birth-and-death evolution of multigene families revisited. *Genome Dyn.* **7**, 170–196.

Elkina M. A., Erkenov T. A. and Glazko V. I. 2015 Mobile genetic elements as a tool for the analysis of genetic differentiation of varieties of cultivated plants and breeds of farm animals. *IJRSR* **6**, 5893–5900.

Esnault C., Maestre J. and Heidmann T. 2000 Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**, 363–367.

Felsenstein J. 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376.

Feschotte C. 2008 Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405.

Gao X., Havecker E. R., Baranov P. V., Atkins J. F. and Voytas D. F. 2003 Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA* **9**, 1422–1430.

Gelfand Y., Rodriguez A. and Benson G. 2007 TRDB—the tandem repeats database. *Nucleic Acids Res.* **35**, D80–D87.

Gozukirmizi N., Yilmaz S., Marakli S. and Temel A. 2015 Retrotransposon-based molecular markers; tools for variation analyses in plants analysis in plants. In *Applications of molecular markers in plant genome analysis and breeding* (ed. K. Taski-Ajdukovic), pp. 19–45, Research Signpost, Kerala.

Gozukirmizi N., Temel A., Marakli S. and Yilmaz S. 2016 Transposon activity in plant genomes. In *Plant omics: trends and applications* (eds. K. R. Hakeem, H. Tombuloğlu and G. Tombuloğlu), pp. 83–108. Springer-Verlag: Springer International Publishing Switzerland.

Guliyev M., Yilmaz S., Sahin K., Marakli S. and Gozukirmizi N. 2013 Human endogenous retrovirus H (Herv-H) genome insertion variations in humans. *Mol. Med. Rep.* **7**, 1305–1309.

Gregory S. G., Barlow K. F., McLay K. E., Kaul R., Swarbreck D., Dunham A. *et al*. 2006 The DNA sequence and biological annotation of human chromosome 1. *Nature* **441**, 315–321.

Havecker E. R., Gao X. and Voytas D. F. 2005 The sireviruses, a plant-specific lineage of the Ty1/copia retrotransposons, interact with a family of proteins related to dynein light chain. *Plant Physiol.* **139**, 857–868.

Hayward A. 2017 Origin of the retroviruses: when, where, and how? *Curr Opin Virol.* **25**, 23–27.

Hřibová E., Neumann P., Matsumoto T., Roux N., Macas J. and Doležel J. 2010 Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biol.* **10**, 204.

Heras J., Dominguez C., Mata E., Pascual V., Lozano C., Torres C. *et al*. 2015 GelJ – a tool for analyzing DNA fingerprint gel images. *BMC Bioinformatics* **16**, 270.

Holligan D., Zhang X. Y., Jiang N., Pritham E. J. and Wessler S. R. 2006 The transposable element landscape of the model legume *Lotus japonicas*. *Genetics* **174**, 2215–2228.

Ivancevic A. M., Walsh A. M., Kortschak R. D. and Adelson D. L. 2013 Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. *BioEssays* **35**, 1071–1082.

Jaccard P. 1908 Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* **44**, 223–270.

Jurka J. 2000 Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420.

Kalendar R., Grob T., Regina M., Suoniemi A. and Schulman A. 1999 IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98**, 704–711.

Karlik E., Gurbuz O., Yildiz Y. and Gozukirmizi N. 2021 Endogenous retrovirus HERV-K6 and HERV-K11 polymorphisms' analyses in head and neck squamous cell carcinoma patients. *Meta Gene* **28**, 100876.

Kazazian Jr H. H., Wong C., Youssoufian H., Scott A. F., Phillips D. G. and Antonarakis S. E. 1988 Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166.

Kojima K. K. 2018 Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob. DNA* **9**, 2.

Kumar S., Stecher G. and Tamura K. 2016 MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.

Laland K. N., Uller T., Feldman M. W., Sterelny K., Müller G. B., Moczek A. *et al*. 2015 The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc Biol Sci.* **282**, 20151019.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J. *et al*. 2001 Initial sequencing and analysis of the human genome. *Nature.* **409**, 860–921.

Laten H. M. and Morris R. O. 1993 SIRE1, a long interspersed repetitive DNA element from soybean with weak sequence similarity to retrotransposons: initial characterization and partial sequence. *Gene* **134**, 153–159.

Laten H. M., Havecker E. R., Farmer L. M. and Voytas D. F. 2003 *SIRE1* an endogenous retrovirus family from *Glycine max*, is highly homogenous and evolutionary young. *Mol. Biol. Evol.* **20**, 1222–1230.

Lynch M. 2007 *The Origins of Genome Architecture.* Sunderland, MA, Sinauer Associates.

Mager D. L. and Stoye J. P. 2015 Mammalian endogenous retroviruses. *Microbiol. Spectr.* **3**, MDNA3-0009-2014.

Malik H. S., Henikoff S. and Eickbush T. H. 2000 Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **10**, 1307–1318.

Mao J., Zhang Q. and Cong Y. S. 2021 Human endogenous retroviruses in development and disease. *Comp. Struct. Biotechnol. J.* **2**, 5978–5986.

Marchetto M. C. N., Narvaiza I., Denli A. M., Benner C., Lazzarini T. A., Nathanson J. *et al*. 2013 Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**, 525–529.

Martin W. F. 2017 Too much eukaryote LGT. *BioEssays* **39**, 1700115.

Martín Sanz A., Gilsanz Gonzales S., Hasan Syed N., Jose Suso M., Caminoero Saldana C. and Flavell A. J. 2007 Genetic diversity analysis in *Vicia* species using retrotransposon-based SSAP markers. *Mol. Genet. Genomics* **278**, 433–441.

McCarthy E. M., Liu J., Lizhi G. and McDonald J. F. 2002 Long terminal repeat retrotransposons of *Oryza sativa*. *Genome Biol.* **3**, 1–11.

Mills R. E., Bennett E. A., Iskow R. C. and Devine S. E. 2007 Which transposable elements are active in the human genome? *Trends Genet.* **23**, 183–191.

Moelling K. and Broecker F. 2019 Viruses and evolution - viruses first? A personal perspective. *Front. Microbiol.* **10**, 523.

Revised Proof

Mun S., Lee J., Kim Y. J., Kim H. S. and Han K. 2014 Chimpanzee-specific endogenous retrovirus generates genomic variations in the chimpanzee genome. *PLoS One* **9**, e101195.

Murphy W. J., Frönicke L., O'Brien S. J. and Stanyon R. 2003 The origin of human chromosome 1 and its homologs in placental mammals. *Genome Res.* **13**, 1880–1888.

Ostertag E. M., Goodier J. L., Zhang Y. and Kazazian Jr H. H. 2003 SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **73**, 1444–1451.

Peccoud J., Cordaux R. and Gilbert C. 2018 Analyzing Horizontal Transfer of Transposable Elements on a Large Scale: Challenges and Prospects. *BioEssays* **40**, 1–8.

Peterson-Burch B. D. and Voytas D. F. 2002 Genes of the Pseudoviridae (*Ty1*/*copia* retrotransposons). *Mol. Biol. Evol.* **19**, 1832–1845.

Popesco M. C., Maclaren E. J., Hopkins J., Dumas L., Cox M., Meltesen L. *et al*. 2006 Human lineage specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313**, 1304–1307.

Ramsay L., Marchetto M. C., Caron M., Chen S. H., Busche S., Kwan T. *et al*. 2017 Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genomics* **18**, 214.

Rebollo R., Romanish M. T. and Mager D. L. 2012 Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46**, 21–42.

Redon R., Ishikawa S., Fitch K. R., Feuk L., Perry G. H., Andrews T. D. *et al*. 2006 Global variation in copy number in the human genome. *Nature* **444**, 444–454.

Schaack S., Gilbert C. and Feschotte C. 2010 Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**, 537–546.

Schulman A. H. and Kalendar R. A. 2005 Movable feast: diverse retrotransposons and their contribution to barley genome dynamics. *Cytogenet. Genome Res.* **110**, 598–600.

Skern-Mauritzen R. and Mikkelsen T. N. 2021 The information continuum model of evolution. *Biosystems* **209**, 104510.

Sorek R., Ast G. and Graur D. 2002 Alu-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067.

Sotero-Caio C. G., Platt R. N., Suh A. and Ray D. A. 2017 Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol Evol* **9**, 161–177.

Soucy S. M., Huang J. and Gogarten J. P. 2015 Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482.

Tamura K. and Nei M. 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526.

Tuzun E., Sharp A. J., Bailey J. A., Kaul R., Morrison V. A., Pertz L. M. *et al*. 2005 Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732.

Villarreal L. P. and Witzany G. 2010 Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* **262**, 698–710.

Viviani A., Ventimiglia M., Fambrini M., Vangelisti A., Mascagni F., Pugliesi C. *et al*. 2021 Impact of transposable elements on the evolution of complex living systems and their epigenetic control. *Biosystems* **210**, 104566.

Wallau G. L., Ortiz M. F. and Loreto E. L. S. 2012 Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. *Genome Biol. Evol.* **4**, 689–699.

Wang L., Norris E. T. and Jordan I. K. 2017 Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. *Front. Microbiol.* **8**, 1418.

Wang L. and Jordan I. K. 2018 Transposable element activity, genome regulation and human health. *Curr. Opin. Genet. Dev.* **49**, 25–33.

Warburton P. E., Hasson D., Guillem F., Lescale C., Jin X. and Abrusan G. 2008 Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics* **9**, 533.

Ward M., Zhao S., Luo K., Pavlovic B., Karimi M. M., Stephens M. and Gilad Y. 2018 Silencing of transposable elements may not be a major driver of regulatory evolution in primate induced pluripotent stem cells. *eLife* **7**, e33084.

Weber B., Wenke T., Frommel U., Schmidt T. and Heitkam T. 2010 The Ty1-copia families SALIRE and Cotzilla populating the *Beta vulgaris* genome show remarkable differences in abundance, chromosomal distribution, and age. *Chromosome Res.* **18**, 247–263.

Wei W., Gilbert N., Ooi S. L., Lawler J. F., Ostertag E. M., Kazazian H. H. *et al*. 2001 Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol.* **21**, 1429–1439.

Wildschutte J. H., Williams Z. H., Montesion M., Subramanian R. P., Kidd J. M. and Coffin J. M. 2016 Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci. USA* **113**, E2326–E2334.

Wolf Y. J. and Koonin E. V. 2013 Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837.

Wright S. I., Le Q. H., Schoen D. J. and Bureau T. E. 2001 Population dynamics of an Ac-like transposable element in self- and crosspollinating *Arabidopsis*. *Genetics* **158**, 1279–1288.

Vandepoele K., Van Roy N., Staes K., Speleman F. and Van Roy F. A. 2005 Novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol. Biol. Evol.* **22**, 2265–2274.

Vandepoele K. and van Roy F. 2007 Insertion of an HERV(K) LTR in the intron of NBPF3 is not required for its transcriptional activity. *Virology* **362**, 1–5.

Vandepoele K., Andries V., Van Roy N., Staes K., Vandesompele J., Laureys G. *et al*. 2008 A constitutional translocation t(1;17)(p36.2;q11.2) in a neuroblastoma patient disrupts the human NBPF1 and ACCN1 genes. *PLoS ONE* **3**, e2207.

Vandepoele K., Andries V. and Van Roy F. 2009 The NBPF1 promoter has been recruited from the unrelated EVI5 gene before simian radiation. *Mol. Biol. Evol.* **26**, 1321–1332.

Yohn C. T., Jiang Z., McGrath S. D., Hayden K. E., Khaitovich P., Johnson M. E. *et al*. 2005 Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol.* **3**, e110.

Yuzbasioglu G., Yilmaz S., Marakli S. and Gozukirmizi N. 2016 Analysis of *Hopi/Osr27* and *Houba/Tos5/Osr13* retrotransposons in rice. *Biotechnol. Biotechnol. Equip.* **30**, 213–218.

Zimmerly S. and Semper C. 2015 Evolution of group II introns. *Mob. DNA.* **6**, 7.

Zhuo X. and Feschotte C. 2015 Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathog.* **11**, e1005279.

Corresponding editor: Durgadas P. Kasbekar