

Article - Biological and Applied Sciences

Au Family Short Retroposons Contribute to Transcriptional and Phenotypic Diversity in Tomato (*Solanaceae*)

Mauro Grabiele^{1,2*}

<https://orcid.org/0000-0001-7386-4924>

Patricia Mabel Aguilera^{1,2}

<https://orcid.org/0000-0002-5063-9350>

¹National Scientific and Technical Research Council and National University of Misiones, Institute of Subtropical Biology, Posadas, Misiones, Argentina; ²National University of Misiones, Misiones Institute of Biotechnology, Posadas, Misiones, Argentina.

Editor-in-Chief: Alexandre Rasi Aoki
Associate Editor: Alexandre Rasi Aoki

Received: 07-Mar-2021; Accepted: 24-Nov-2021.

*Correspondence: maurograbiele@conicet.gov.ar; Tel.: +54-0376-4435099 (M.G.).

HIGHLIGHTS

- Tomato genome comprises ca. 670 Au SINE copies and >80% associates to genes.
- Tomato Au SINEs are transcribed inside genic circRNAs and participate of mature mRNA.
- Tomato Au sequences translation originate novel protein domains and locus protein variants.
- Biased survival of Au SINEs at 486 tomato protein coding genes appears an adaptive feature.

Abstract: Here we report the current gene impact of the Au family of SINEs in tomato. The genome of *Solanum lycopersicum* 'Heinz 1706' SL3.0 -NCBI annotation release 103- was reference searched and the Au profile was characterized in-depth. Tomato genome comprises ca. 670 Au copies, of entire length -18.5%- or truncated, randomly inserted and eroded, forming three well supported (>80%) super clusters which disperse along the 12 chromosomes mirroring the subtelomeric gene distribution bias of the species. In tomato, the Au clade is largely localized at protein coding genes -69.5% introns, 7.8% 3UTRs, 2.1% 5UTRs, 1.2% CDSs- followed by genomic copies -18.3%-, long non coding RNA genes -1.4%- and pseudogenes -0.8%-. The 419 tomato genes harboring intronic Au are diverse, weakly associated considering biological processes and molecular functions, but include important traits such as stress response, hormone response or phenotype plasticity. Au was found to be transcribed inside circular RNAs derived from 12 genic loci. Exonic Au affect the transcriptional and/or translational profiles of 67 tomato genes, including biological/agronomical important ones, contributing to UTR length and composition, UTR transcript variants, CDS boundary definitions, protein domains and variants. We propose that biased survival of Au in tomato genes is an adaptive feature.

Keywords: SINE Au; genome; introns; circRNAs; exonization; UTRs; CDSs.

INTRODUCTION

The eukaryote genome is mainly composed of transposable elements such as transposons and retrotransposons that moves as DNA or RNA molecules, respectively [1]. These elements constitute a source of biodiversity since they can modify the genotype and phenotype, i.e. during developmental and reproductive stages, leading to genomic stability and evolution [2-6].

Particularly, the short transposable elements SINEs (*Short Interspersed Nuclear Elements*) are a group of non-autonomous retroelements, 70-500 nt in length, which holds a characteristic structure (5' head, body, 3' tail) and its retrotransposition depends on proteins coded by a LINE (*Long Interspersed Nuclear Element*) partner [7-10]. The 5' head of SINEs originates in transference RNAs (tRNAs), 7SL RNA or 5S RNA, and holds the internal promoters A and B necessary to transcription of the entire element by RNA polymerase III (pol-III) [11], but also can derive from small nuclear RNAs (snRNAs) [12]. The body of the SINE is of unknown origin and variable in length, aiding to delineate the SINE families. In contrast, the 3' tail of SINEs is similar to LINEs presumed to be derived -controversial in plants [13]- and also is variable in sequence and length, coulding end in A-rich, AT-rich or microsatellite (i.e. CAn, TTGn) stretches or consecutive Ts (poly-T tail), all of them pol-III transcription termination sites. The insertion of SINEs via LINEs machinery generates novel SINE genomic copies, by amplification without loss of the mother copy, in addition to target sites duplication (TSDs) [13,14].

SINEs affect the eukaryote genome at diverse levels such as causing expansion, insertional mutations at genes -coulding be exonized- and their flanking regions, unequal crossing over mediated deletions/duplications -fostering the emergence of novel genes-, and gene silencing mediated large scale heterochromatinization [7,15-18]. SINEs can regulate gene activity as enhancers/silencers of contiguous genes, by sequestering pol-II transcription factors via their hairpin structure, or by affecting alternative splicing patterns [19-21]. In addition, SINEs can regulate gene transcription and traduction during biotic/abiotic stresses or during development acting as modulators of the small interference RNAs (siRNAs) pathway [22,23]. Particularly in tomato, Quadrana and coauthors [24] reported the case of a SINE (SINE1-SO) which insertion at the promoter region of the gene *VTE3(1)* affects the fruit content of vitamin E.

Au family members of SINEs were first identified in the intron of the acetyl CoA carboxylase gene of *Aegilops umbellulata* Zhuk. and are common to Spermatophyta which last common ancestor is 175 million of years [25-28]. Au comprises ca. 180 nt in length, its 5' head is derived from tRNA and the 3' tail ends with poly-T [26]. This last feature and the absence of poly-T tail LINE partners in plants suggest that the genomic copies of Au are inactive, however the retrotransposition mode of Au has to be discovered yet before concluding that it is a fossil clade of SINEs. Recently, diverse SINE families including Au were found to be deeply associated with genes in Solanaceae genomes [29] and detailed analysis of the wheat transcriptome revealed several mature splice variants of protein-coding genes that carry Au elements [30], suggesting that Au family may play a role in transcriptome and phenotype diversity in plants.

In this sense, and with the aim to reveal the current gene impact of Au members in tomato, we carried out a detailed molecular analysis of this family in the genome of *Solanum lycopersicum* L. 'Heinz 1706'. It is expected that the novel SINE data could be useful to genetics of tomato and other cultivated Solanaceae.

MATERIAL AND METHODS

Tomato genome assembly version SL3.0 and corresponding data from annotation release 103 were downloaded from the National Center for Biotechnology Information (NCBI) FTP site https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/4081/103/GCF_000188115.4_SL3.0/ and used to build a local database on the software Geneious 11 (Biomatters Ltd.). To further characterize the Au SINE family profile in the tomato genome, in house Blastn reference searches (cut off e-05) were conducted using the consensus Au sequence of *S. lycopersicum* SL2.5 [29], annotated via the original Au sequence of *A. umbellulata* [26], as query. Overall obtained genome nucleotide hits -annotated all through with gene/genomic GFF3 markers- were further mapped against the query Au sequence using the Geneious mapper tool at default values to check and finally annotate the corresponding Au features on them. The mapping matrix was used to build an approximately maximum likelihood (ML) phylogenetic tree of Au sequences via FastTree 2.1.5, which estimates splits reliability by Shimodaira-Hasegawa test and 1,000 default resamples, using the substitution model General Time Reversible (GTR) with a gamma -20- distribution of rates of evolution among sites [31]. Tomato Au hits were so classified in major clusters according to tree topology, and this information added to chromosomal coordinates and nucleotide sense of Au was annotated onto sequences via GFF3 format. These Au sequences and their annotated features were directly mapped against tomato

chromosomes via the GFF3 import protocol of Geneious. To avoid mapping errors, an additional mapping of Au sequences was performed via a nucleotide search at 100% similarity and index length of 10 onto tomato chromosomes by means of the Annotate & Predict tool of Geneious and finally both mapping strategies results were visually compared and checked for consistency. At this point, mapped Au sequences were classified according to target locus into main categories such as genomic, protein coding gene, long non coding RNA (lncRNA) gene or pseudogene, and those genic Au were further classified according to targeted internal structures namely intron, exon, coding sequences (CDS) and untranslated sequences at 5' or 3' (5UTR or 3UTR, respectively).

Overall tomato described 1976 circular RNA sequences (circRNAs) were downloaded from the Plant Circular RNA Database site <http://ibi.zju.edu.cn/plantcircbase/> (v4_1976sly_genomic_seq). These sequences were scanned through the annotated Au consensus sequence of tomato [29] during in house Blastn searches (cut off e-05). Full length of circRNAs reached by Au were directly mapped onto tomato chromosomes as explained above.

Selected Au hits on genes and circRNAs of tomato were further curated in length to reach polyT and TSDs features. Paralogy and orthology analysis of Au containing genes of tomato were performed via in house BlastP searches (cut off e-05) in the tomato genome and those of potato (*Solanum tuberosum* L. clone DM1-3 516R44 3.0, NCBI annotation release 101) and chili pepper (*Capsicum annuum* L. 'Zunla-1' 1.0, NCBI annotation release 100), respectively, then properly checked at Ensembl Plants Database [32]. Nucleotide alignments of Au containing genes of tomato with respective paralogs and orthologs were performed by Mafft v7.308 at default values. Proteins were aligned via MUSCLE 3.8.425 at default values and further annotated at Pfam [33]. Splice variants of Au containing genes of tomato at CDSs and UTRs were validated at NCBI Gene Database [34] by curated RNA-seq samples alignments supported introns of tomato annotation release 103 and then compared in abundance through the equally supported number of spliced reads among corresponding regions. Gene ontology (GO) enrichment analysis for biological process (BP) and molecular function (MF) were performed at AmiGO 2 [35], using the annotated tomato genes as reference at each category (BP: 8246/34637; MF: 8783/34637) and Fisher's exact significance test with the Bonferroni correction ($P < 0.05$). Linear regression analysis among variables considering Pearson's correlation coefficient (R) and statistical graphs were conducted in Microsoft Excel 2010.

RESULTS AND DISCUSSION

General features of the Au SINE family in tomato

The similarity-based search approach in the tomato genome SL3.0 revealed 672 Au hits between e-61 to e-05 values, displaying 65% average identity and 112 nt mean length (189 to 33 nt). Similarity-based strategies in former assembly versions of the tomato genome found 701 (SL2.4; [36]) and 604 (SL2.5; [29]) corresponding Au hits. Exclusively 124 out of 672 Au hits represent full-length SINEs while the rest are truncated copies, comparable to the 107 complete sequences found via the ab initio approach through Sine-Finder at scaffolding level of Wenke and coauthors [10], before the assembly release 1.0 of the tomato genome [37]. Those 672 Au copies dispersed fairly in number along different length categories (Figure 1; SINE_Au copies length distribution) and also distributed equally along the consensus sequence (Figure 1A; Coverage) which may be accounted for the typical stochastic erosive processes after genome integration of transposable elements [1], pronounced as expected in largest sequences (189-180 nt category). The low quantity of shortest elements (Figure 1; SINE_Au copies length distribution; 49-30 nt category) was also expected for the resolution threshold employed here (e-05). On the other hand, point variation in Au copies is biased towards depletion of CG nucleotides *sensu lato* (Figure 1A; Identity and Sequence Logo), which constitute target sites to RNA-directed methylation involved in gene silencing and heterochromatin formation [38], hence reducing the chance of the latter phenomena. In addition, the number of Au elements showing a complete tRNA non related region (251) surpasses to those with an entire tRNA related region (204), which is expected for the usual abortive retrotransposition that originates 5' truncated copies of SINEs/LINEs [10,39]. Smallest Au copies being truncated at 5' during retrotransposition were found at the 3UTR of the 4_LOC101258246 gene and one intron of the 7_LOC101255512 gene, with 55 and 51 nt in length including polyT tail and TSDs, respectively (Figure 1B).

Overall Au elements distributed throughout the 12 tomato chromosomes (Table 1), except for a genomic hit at an unplaced genomic scaffold (SL3.0 SL3.00SC0000087). In general terms, the number of Au elements per chromosome do not correspond well with its size ($R=0.45$) or gene content ($R=0.23$) (Table 1; Figure 1) but with a randomized pattern of insertion of copies. Chromosome 2 displays the lowest Au copy number (25;

Table 1; Figure 1) and SINEs are almost restricted to the large arm (Figure 2A), facts probably associated to a small chromosome size (56 Mbp) and an entire short arm being occupied by heterochromatin and the nucleolar organizer region (NOR; [37]).

Overall Au SINE of tomato can be grouped into three well supported (>80%) Super Clusters of sequences (I-III; Figure 1C), which is consistent with the “Master Gene” evolutive model of SINEs in that few functional copies (three in this case) colonize the genome, in contrast with the “Transposon or Multiple genes” model [26,40].

Table 1. Number and distribution of SINE_Au copies in the tomato genome.

Chr	Mbp	Total genes	Gene density	Total Au	Au density	Gene/Au ratio	Genomic	protein coding genes				lnc RNA genes		pseudo genes	
								in	CDS	5UTR	3UTR	ex	in	ex	in
1	98.5	3701	37.6	63	0.6	58.8	12	46	-	3	2	-	1	-	-
2	56.0	2937	52.4	25	0.4	117.5	1	20	-	2	2	-	-	-	1
3	72.3	2932	40.6	55	0.8	53.3	7	38	-	3	5	1	1	-	-
4	66.6	2484	37.3	72	1.1	34.5	9	54	3	-	5	-	-	1	-
5	66.7	2031	30.4	71	1.0	29.9	9	48	2	-	12	-	-	-	-
6	49.8	2581	51.8	42	0.8	63.0	6	28	-	1	4	-	2	-	-
7	68.2	2246	32.9	59	0.9	38.1	17	36	2	1	2	-	-	-	1
8	66.0	2159	32.7	45	0.7	48.0	5	33	1	-	3	2	1	-	-
9	72.9	2122	29.1	58	0.8	36.6	21	33	-	1	2	-	1	-	-
10	65.6	2176	33.2	51	0.8	42.7	8	39	-	-	3	-	-	-	1
11	56.6	1933	34.2	59	1.0	32.8	11	43	-	-	4	-	-	-	1
12	68.1	2101	30.9	72	1.1	28.8	16	45	-	3	8	-	-	-	-

Chr: chromosomes. Mbp: megabase pairs. in: intron. ex: exon.

Current distribution of the Au SINE clade in the tomato genome is biased to intronic regions of protein coding genes (69.5%), followed by genomic copies (18.3%) and 3UTR sequences (7.8%) (Figure 1; Table 1). Au elements are also present at 5UTRs (2.1%) and CDSs (1.2%) of protein coding genes, added to lncRNA genes (1.4%) and pseudogenes (0.8%) (Figure 1; Table 1). Overall tomato SINEs appeared associated to genes (51%) and genomic regions also (48%) [29], which compared to the analysis of Au performed here revealed that this family contributes largely to the total genic SINE fraction. Tomato chromosomes harbor a higher density of genes at subtelomeric regions whilst centromeric and pericentromeric regions are almost depleted of genes but largely formed by heterochromatin originated in Copia and Gypsy LTR-retrotransposons [37,41]. Global mapping of the 671 genomic or genic Au copies in the tomato chromosomes exposed a distribution pattern that follows that of the density of genes (Figure 2A and B). Most parsimonious hypothesis to explain the current Au distribution in tomato is that genome copies amplified as usual via pol-III and inserted randomly at different locations to finally being eliminated at centromeric and pericentromeric regions of chromosomes by the activity of major representatives of LTR-retrotransposons. At the same time, the current intimate association of Au elements to gene structures (>80%) and the CG decrease bias of inserted Au copies that can surpass gene silencing could be explained in terms of an adaptive advantage to tomato, in which the performance of invaded and/or adjacent genes is affected.

Au SINEs in spliceosomal introns of genes of tomato

Au elements at introns were biased to the Super Cluster III of sequences (46.5%), like those Au SINEs of genomic regions (51.6%) (Figure 2), and linked to 419 protein coding genes, six lncRNA genes and four pseudogenes (Table 1; Table S1, available at <<http://dx.doi.org/10.17632/x4s5j9cs96.1>>). Tomato assembled chromosomes hold more than five thousand annotated lncRNA genes, and ca. 7% originated in LTR retrotransposons [42], in contrast to Au contribution. Tomato genes harboring intronic Au are diverse and embrace important agronomic traits such as stress response (TFT4, LOC101246054, LOC101243790, LOC101260189, LOC101260794, LOC104645214, LOC101258771, SAP13), hormone response (ARF12, LOC100191131, ARF9), or phenotype plasticity mediated by transcription/translation/splice factors (LOC101248963, LOC101251993, LOC101254825, LOC101252686, LOC101262193, LOC101249451, LOC101248589, LOC101262521, LOC101261183, LOC101259473). In addition, an Au SINE was found in antisense direction at the intron of the transposase of a MuDR family member (LOC101247243).

GO enrichment analysis for biological process of the Au containing introns of protein coding genes found 153 genes distributed in 103 categories, with only 26 genes significantly enriched (2.5 fold; $P=3.5e-02$) to phosphorus metabolism, and 56 (2.0 fold; $P=1.4e-03$) to organonitrogen compound metabolism, both

constituting general pathways. Additionally, GO enrichment analysis for molecular function found 168 genes in 101 categories, with only 17 genes significantly enriched (3.8 fold; $P=5.1e-03$) to protein serine/threonine kinase activity, which is a wide-ranging function. Hence, GO enrichment analysis did not find strong association among tomato genes with intronic Au, which a priori is consistent with the model of random insertion of SINE copies, as explained above. On this regard, stringent megablast searches (word size 28; cut off $e-100$) in the pool of Au containing introns found solely 2 paralogous gene relationships among data, i.e. at the homeobox-leucine zipper protein HDG11-like genes 7_LOC101265456 and 7_LOC101255311, and the alpha-aminoadipic semialdehyde synthase gene 7_LOC101261722 and its corresponding pseudogene 7_LOC101256395. This very low frequency of paralogs carrying an equivalent Au in introns that can be involved in the same processes or with similar functions may be the result of events such as paralogues split before Au insertion combined to typical accumulation of polymorphisms in introns at paralog rather than ortholog level [43,44].

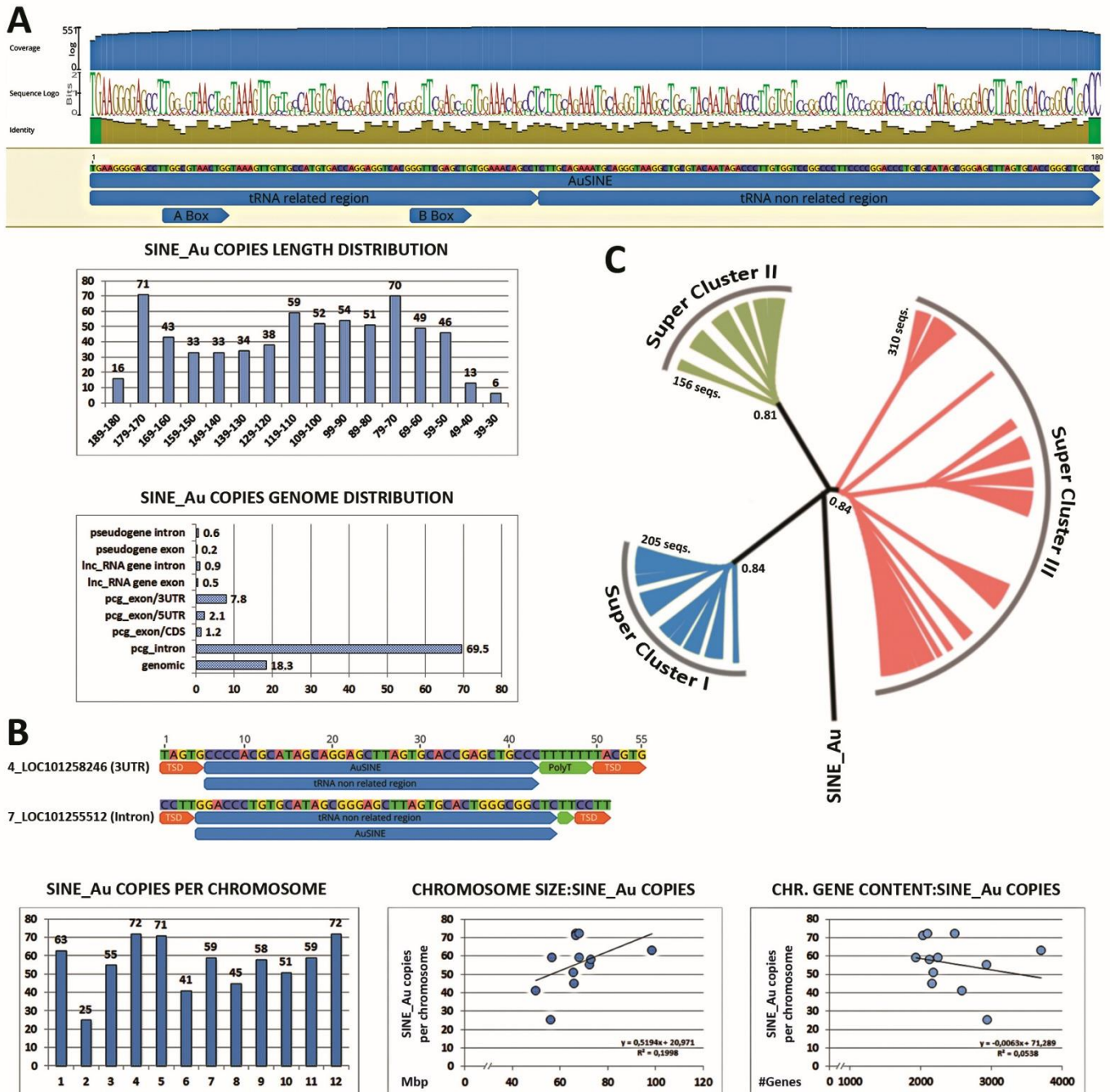


Figure 1. Characterization of the Au SINE family in tomato. A) Annotated consensus Au sequence and respective identity, sequence logo and coverage graphs of the 672 Au hits. B) Annotated smallest Au elements. C) ML phylogenetic tree of overall Au sequences with respective support values and rooted with the consensus SINE_Au. Bar and linear regression graphs consider all Au copies.

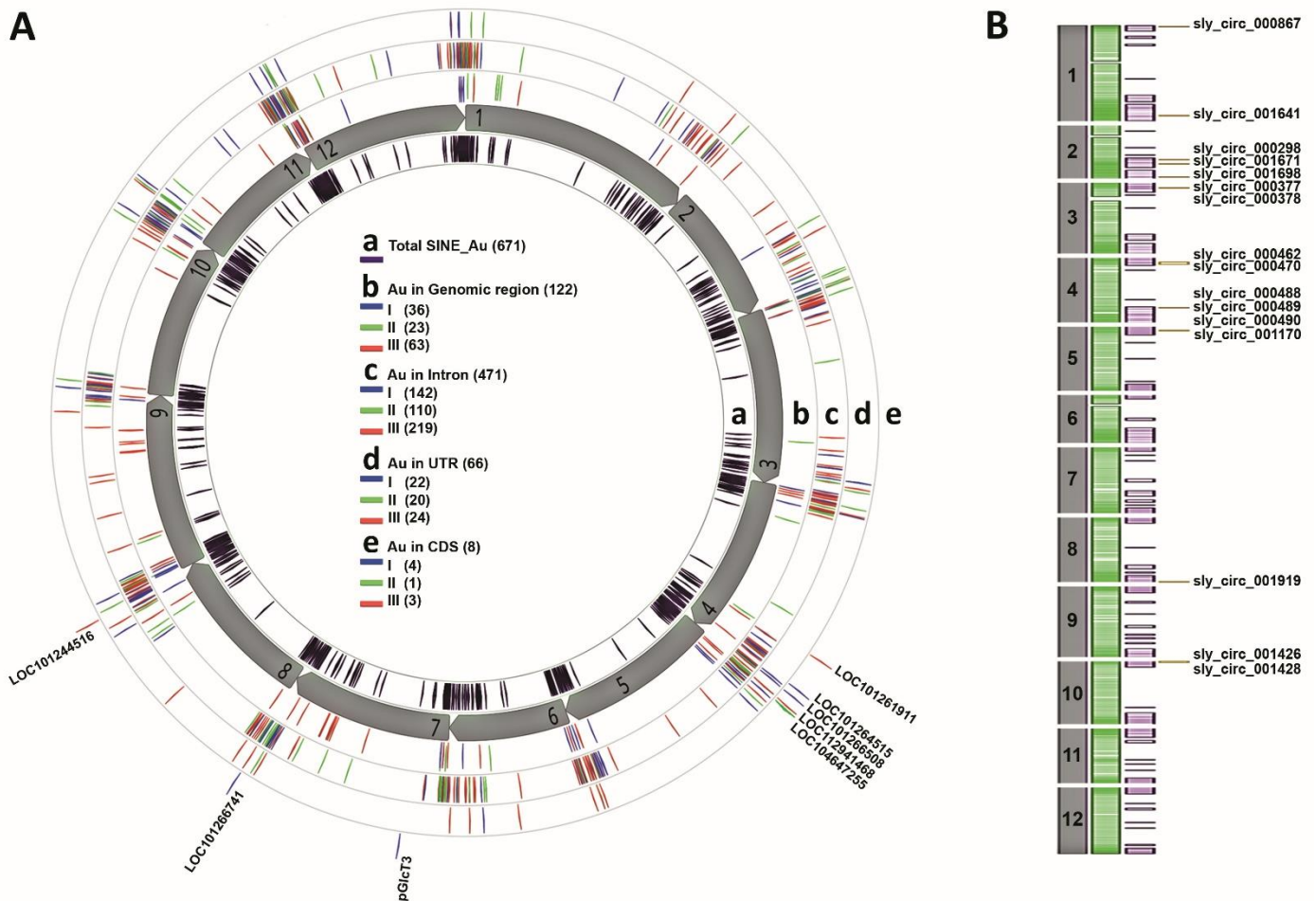


Figure 2. Characterization of the Au SINE family in tomato. A) Chromosomal map of overall Au elements, general one (a) and according to major categories (b-e) and clades (I-III). Gene locus names of Au SINEs containing CDSs (e) are pointed out in the map. B) Chromosomal densities map correspondence of genes (green) and overall Au SINEs (violet); high density regions are denoted by intense colours. Au containing circRNAs derived from gene locus are mapped onto chromosomes also (yellow).

The intron sequence is involved in crucial biological features such as the regulation of alternative splicing, the positive regulation of gene expression, the regulation of nonsense-mediated decay, the control of mRNA transport or chromatin assembly and the genome stability [45-48]. In addition, intron sequences are a source of various types of non coding RNAs (ncRNAs) in genomes [49]. Particularly, and beyond their abundance, the outstanding feature of the tomato spliceosomal introns harboring Au elements is that their sequences became larger after SINE integration, among ca. 50 to 200 nt in length. In this sense, there is increasing evidence that intron length contributes directly to the control of gene expression such that evolutionary old genes that transcribe early in development and those involved in rapid biological responses tend to have shorter introns than tissue specific genes [50-53]. Further, Zhang and coauthors [54] hypothesized that introns could achieve gene expression and regulation by interacting with corresponding mRNA, such that intron length and mRNA sequences co-evolve to successfully perform their biological functions. Also it was postulated that long introns in higher eukaryotes constitute a good reservoir of proto-genes, which transcription and translation might provide adaptive potential to cells in different physiological environments [55]. A remarkable feature is that intron length affects the recombination rate and the efficiency of natural selection in finite populations [56]. In this sense, genes with longer introns are under weaker Hill-Robertson interference by increasing recombination, which ultimately enhances the chance for two favorable alleles at linked loci to be located together [47]. Accordingly, overall evidence supports the hypothesis that current biased survival of Au SINEs at spliceosomal introns of tomato genes constitutes an adaptive feature.

Another subject to be considered regarding the adaptive value of Au sequences at introns of tomato is the phenomenon of intron retention, the most frequent form of alternative splicing in plants in that introns are included in mature mRNA transcripts, able to generate alternative protein isoforms with novel functions and playing key roles in normal development and under stress conditions [57]. Though it is expected that pol-II

transcripts with spliceosomal intronic SINEs suffer post splicing degradation according to the canonical pathway [58], the escape from this via and a post splicing function was also reported for pol-II transcribed spliceosomal introns, such that they can act as precursors of diverse RNAs, -i.e. micro RNAs derived from introns (mirtrons)-, modify the chromatin conformation, regulate the gene expression and splicing at different levels or act during development [59,60]. In this sense, Buckley and coauthors [61] reported the case of an intronic retained SINE in the cytoplasm of dendrites, which modify the physiological function of the proteins coded by the SINE containing gene. Future studies on the occurrence of intron retention in tomato will shed light on the contribution of the Au containing introns to its transcripts profile considering its abundance and the remarkable feature that transposable elements insertion at introns were found to enhance intron retention events [62].

Au SINEs in tomato circular RNAs derived from genic loci

Regarding the similarity-based search approach in the informed 1976 tomato circRNAs, 16 Au hits (0.8%) were exposed showing 64 to 88% pairwise identity to the Au consensus sequence. Those 180 to 48 nt in length Au hits constitute complete (54%) and 5' truncated copies (38%), holding TSDs and polyT tail, added to a 3' truncated copy (8%) with recognized TSDs. Further, the Au containing circRNAs (675 to 27105 nt in length), which mapped onto tomato chromosomes 1 to 5, 8 and 10 (Figure 2B), embraced multiple introns and exons of 12 different protein coding genes and a unique lncRNA gene (8_LOC109120968) (Figure 3). Hence, Au elements linked to circRNAs reside at introns (75%), UTRs (17%) or CDSs (8%) from those protein coding genes while that of the lncRNA gene is exonic (Figure 3). Particularly, three different size circRNAs, i.e. sly_circ_000488 (9750 nt), sly_circ_000489 (4250 nt) and sly_circ_000489 (2013 nt) mapped to the crossover junction endonuclease EME1B gene LOC101261911 and all of them carry the same full length Au SINE which is fundamental to the expression of an alternative CDS splice variant of the locus (see below).

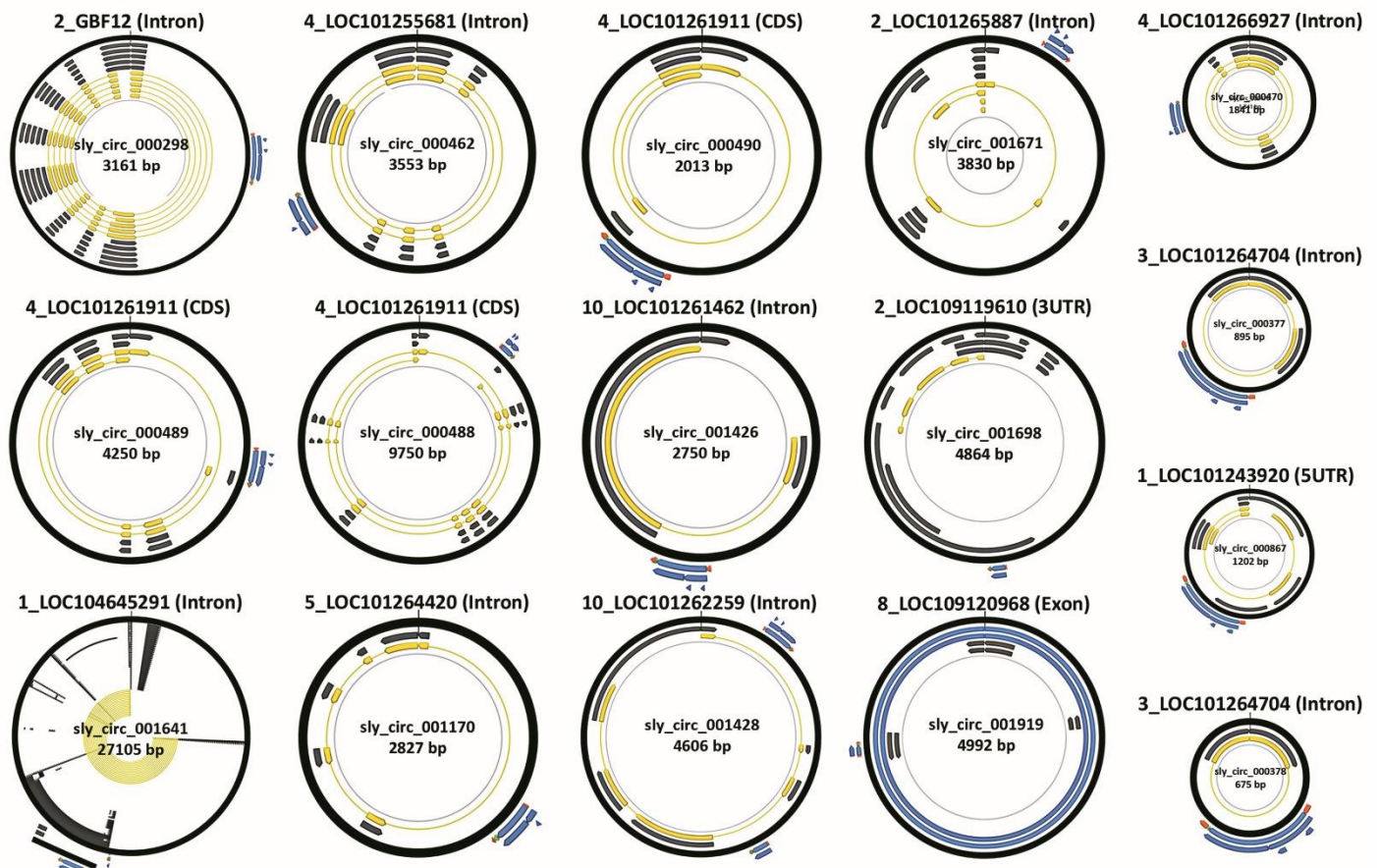


Figure 3. Annotated Au containing circular RNA sequences derived from protein coding genes and a lncRNA gene (LOC109120968). Mapped Au elements are annotated in blue with their respective TSDs (orange) and polyT tail (green). Exons (grey), coding sequences (yellow) and the lncRNA gene circularized transcript (blue) are also annotated onto circRNA sequences. A detail of the Au copy at the largest circRNA (sly_circ_001641) is shown below the Figure.

The circular RNA sequences are a class of endogenous ncRNAs able to influence gene expression by modulating the function of regulatory ncRNAs, i.e. microRNAs, and RNA-binding proteins, acting as sponges or scaffolds [63,64]. In plants, including tomato, circRNAs were found to be differentially expressed during biotic or abiotic stresses [65]. Two recent reports in tomato revealed a thousand different circRNAs [66,67] and 1976 sequences are deposited at the Plant Circular RNA Database, though this number constitute an underestimation considering the ca. 30 thousand circRNAs characterized in Arabidopsis to date at the mentioned database. According to the encircled regions, the Au containing circRNAs of tomato can be further characterized as exon-intron circRNAs (EircRNAs), a particular class localized in the nucleus [63]. A significant close association between circRNAs and transposable elements was first described for Alu SINEs in human genes [68]. Most recently, Chen and coauthors [69] reported the linkage between circRNAs and LINE1-like elements of maize genes, remarkably able to modulate transcriptomic and phenotypic variations in this crop. In both cases, retroposons and their reverse complementary sequences flanking circRNAs contribute to circRNAs formation and accumulation. Here we describe the association between circRNAs and Au SINE sequences in tomato genes in which Au elements are contained within circRNAs instead at their flanking regions. Further studies are needed to disclose the real impact of Au containing circRNAs -a novelty in plants- in the physiology of tomato.

Au SINEs in UTRs of transcripts and translated gene sequences of tomato

Au elements at exonic regions were biased to UTRs (85%) rather than CDSs (15%), associated to 67 protein coding genes and contributing in several ways to the transcriptional and translational profiles of tomato (Table 2). In this sense, Au SINE sequences contribute to the length and composition of the 3UTR (79%) and 5UTR (21%) of transcripts derived from 47 diverse protein coding genes, including biological and agronomical important ones such as the late blight resistance protein homolog r1a-3 (1_LOC101257414), the ethylene-responsive transcription factor erf054 (7_LOC101264838), the nuclear transcription factor y subunit a-7-like (12_LOC101247294) and the transcription factor tcp12-like (5_LOC104647293) among others (Figure 4A; Table 2). The UTRs of eukaryotic mRNAs are essential noncoding regulatory elements for post-transcriptional gene expression [70]. Contrasting to 3UTR sequences, 5UTRs are more conserved and contain various regulatory elements such as AUG start codons and internal ribosome entry sites (IRES) upstream of ORFs, involved in the control of translation initiation [71]. At this regard, transposable elements such as LINEs and SINEs were found at 5UTRs causing upstream ORFs (uORFs) that participate in the regulation of the main ORF translation [72]. In agreement, Au SINE sequences that affect to the length and composition of the 5UTR of tomato transcripts derived from four different loci also were found to constitute uORFs, i.e. 1_LOC101257414, 2_LOC101249596, 3_LOC101268270 and 12_LOC101247294 (Table 2). In addition, transposable elements such as MITEs, Helitrons, LINEs and SINEs were also found to contribute to the length and composition of 3UTR sequences in diverse species associated to post-transcriptional and translational levels regulation [62,73-77] which highlight our findings in tomato.

Au SINEs also contribute with novel splice sites, i.e. canonical AG and GU, or non-canonical GG, to UTR transcript variants of seven protein coding genes (Figure 4B; Table 2), including the significative tm-1 protein, which confers resistance to the Tomato Mosaic Virus [78]. Those UTR transcript splice variants of genes were further characterized according to the RNA-seq supported introns and number of spliced reads associated to each gene model data. In this sense, 5UTR regular mRNAs lacking Au SINEs were found to contribute more to total transcripts of the locus than those 5UTR splice variants harboring Au sequences, i.e. 37.6:1 for 1_LOC101256630, 46.3:1 for 1_LOC101243920, 7.8:1 for 2_LOC101252742 and 32.5:1 for 6_5PT3 (Figure 4B). On the other hand, Au SINEs containing 3UTR splice variants are more abundant than regular 3UTR transcripts lacking Au elements of the same locus, i.e. 21.6:1 for 2_TM-1 and 2.5:1 for 12_LOC101257050 (Figure 4B). In any case, splice variants of UTRs of tomato transcripts caused by Au SINE insertions could be valuable at physiological level since the tissue-specific expression of transcripts with alternative spliced UTRs can control protein expression [79].

Au SINE insertions were also found to affect the length and composition of the UTR as well as the CDS boundary definition, via novel start or stop codons sites, of five tomato protein coding genes (Figure 4C; Table 2), features reported for different SINE families in potato too [29]. Furthermore, Au SINE sequences are also translated contributing to protein domains of eight different tomato genes (Figures 4D-E and 5A-C; Table 2). Particularly, paralog loci 5_LOC112941468 and 5_LOC104647255 code for proteins variants which terminal domains are affected by an ancestral Au SINE insertion and subsequent sequence duplication and deletion events (Figure 4D). The oligopeptide transporter 4-like protein of 4_LOC101264515 of tomato exhibits two

splice variants promoted by a novel splice site provided by a terminal Au element insertion (Figure 4E). This way, contrary to the largest length protein variant (XP_004238028), the shortest one carry the Au SINE features at its terminal domain (XP_025886618), but the former is more abundant in relation 60.1:1 (Figure 4E). Regarding the eighth tomato paralog members related to gene 4_LOC101264515, only the latter harbors the mentioned Au element, which is also present at potato ortholog LOC102599682 but absent from the corresponding orthologue of *C. annuum* (Figure 4E).

In addition, the insertion of an Au SINE at gene 4_LOC101261911 of tomato generated novel splice and start codon sites allowing to two transcript variants, in which the regular one contributes more to total expression of the locus than the Au containing transcript in relation 16.5:1 (Figure 5A). Hence, gene 4_LOC101261911 of tomato produce two protein variants of the crossover junction endonuclease EME1B, the shortest one defined by an Au derived N-terminus (Figure 5A). Ortholog genes such as potato 4_LOC102605330 and chili pepper 12_LOC107851234 also share the corresponding Au SINE from the last common ancestor of *Solanum* and *Capsicum* (ca. 20 million years; [80]), but solely in tomato it was exonized and translated, this last prevented particularly in potato by a mutation at the ATG site (Figure 5A). Moreover, gene 8_LOC101244516 of tomato produce three variants of the tpx2 protein, one of them carrying a truncated and antisense Au element at its N-terminus, which defined novel splice and start codon sites at the locus. In addition, the largest regular transcript of the locus is more abundant than the Au containing one in relation 9.1:1 (Figure 5B).

Tomato 8_LOC101244516 and potato ortholog gene LOC1012600111 also share the corresponding Au SINE from their last common ancestor (ca. 7 million years; [80]) however, in potato it can not be translated since its ATG site is mutated (Figure 5B). The corresponding ortholog gene of *C. annuum* -LOC107866940-lacks the truncated and antisense Au element present at *Solanum* members but exhibits another sense and entire length Au copy a few bases upstream from the former (Figure 5B).

Finally, gene 7_PGLCT3 of tomato that codes for the plastidic glucose transporter 2-like protein holds an Au SINE insertion which contributes to a novel splice site at its middle region allowing to two transcript variants, one regular and more abundant in relation 17.2:1 than the other one carrying the Au element (Figure 5C). The tomato paralog gene LOC101250024 lack the Au SINE insertion, which is also absent in the *C. annuum* orthologue LOC107878000 (Figure 5C). However, the Au element is present at the ortholog gene LOC102578833 of potato in which is not translated as in tomato probably due to a mutation in the 3' splice site (Figure 5C).

Exonization and subsequent translation of SINEs was previously reported for Alu elements of primates [81] and recently in the Au family of wheat [30]. According to the former author, SINE insertions are not detrimental due to alternative splicing mechanism, through which the organism can produce the regular protein in addition to the new variant/s encompassing the SINE that ultimately can be beneficial. Regarding the analysis in wheat of Keidar and coauthors [30], Au SINE containing mature transcripts constantly demonstrated lower expression levels than regular transcripts lacking the Au element, the same as observed here in tomato. This biased expression levels among regular and alternative Au exonized transcripts might indicate that the later can provide proteins with new functions, as hypothesized by Schmitz and Brosius [82], potentially useful to the adaptation to changing environments, particularly in a species with limited genetic variation as modern crops including tomato.

CONCLUSION

Tomato genome comprises ca. 670 Au copies, of entire length or mostly truncated -81.5%-, that randomly inserted and suffered erosion, which originated from at least three ancestral SINE master elements. Au copies spread along the 12 chromosomes mirroring the subtelomeric gene distribution bias of tomato. Protein coding genes are the largest reservoir of the Au clade in tomato genome encompassing the 80.6% of copies. Au also colonized other genomic regions -18.3%-, lncRNA genes -1.4%- and pseudogenes -0.8%-. A total of 419 protein coding genes of tomato harbor intronic Au elements, 63 are inhabited by Au copies at UTRs and 8 at CDSs. Appart from being transcribed inside genic circular RNAs, tomato Au SINEs participate of mature mRNA transcripts. Hence, Au affect largely to UTR length and composition of transcripts. In addition, Au contribute with novel splice sites and start or stop codons that outline CDS boundaries and promote UTR and CDS transcript variants at the same locus. Those 5UTR and CDS transcripts variants harboring Au sequences contribute less to the total expression of the locus than regular transcripts, contrary to the 3UTR variants. Translation of Au sequences in tomato originate novel protein domains and locus protein variants. Au contribution to novel gene functions added to Au potential involvement in post-transcriptional and translational levels regulation support the hypothesis that current biased survival of Au

SINEs at 486 tomato protein coding genes, including biological and agronomical important ones, is an adaptive feature.

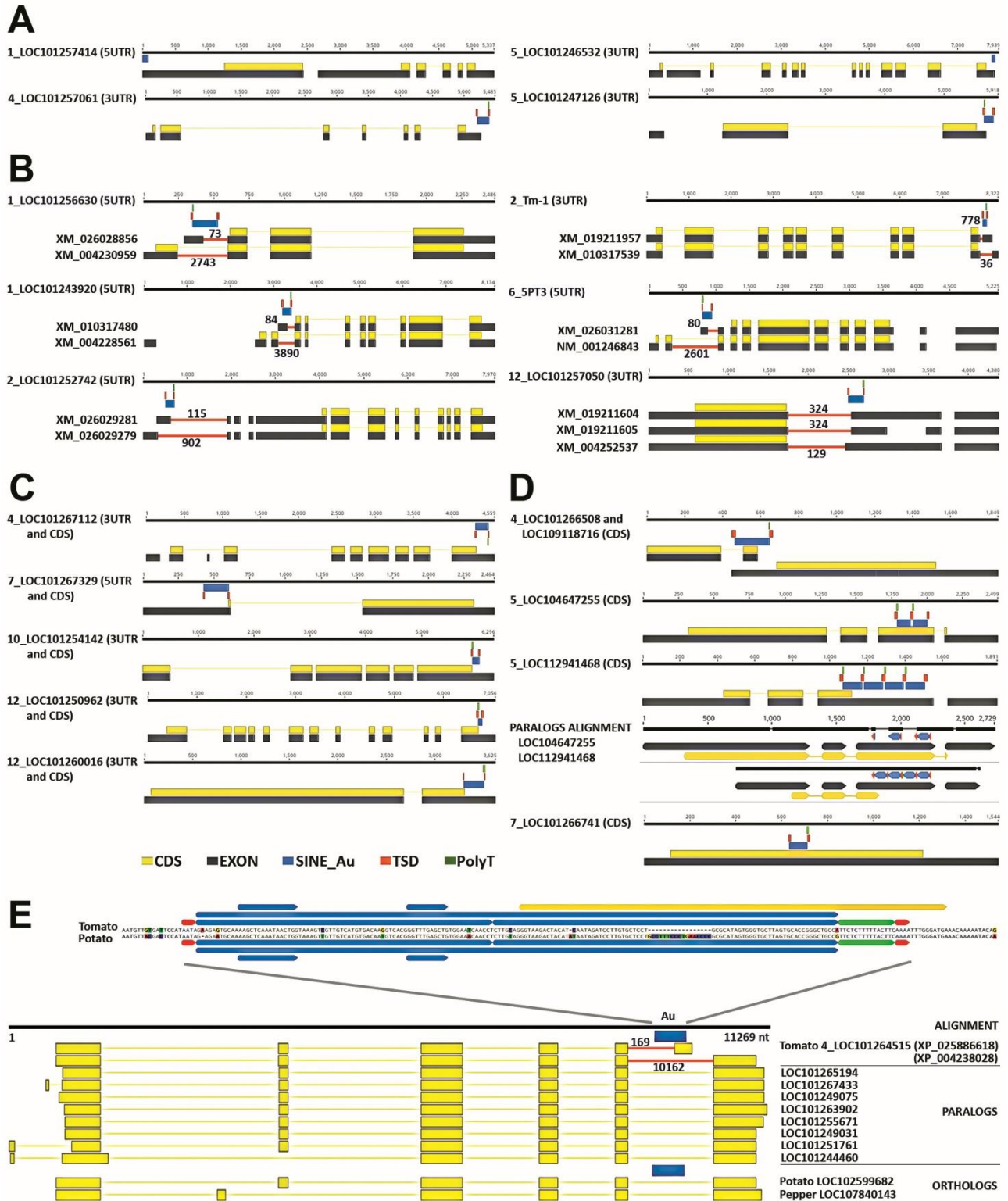


Figure 4. Au SINE annotated tomato gene models showing Au contribution to exonic regions. A) UTR length and composition, including the 5UTR of the late blight resistance protein homolog r1a-3 gene. B) UTR transcript variants via novel splice sites, including the 3UTR of the tm-1 protein gene conferring resistance to TMV. C) CDS boundary definition via novel start or stop codons. D) CDS and protein domains. Note the alignment of two paralogs which shared Au insertion and evolution resulted in proteins variants. E) CDS and protein variants. Alignment of the oligopeptide transporter 4-like protein gene, its paralogs and orthologs; Au SINE alignment region is amplified. Compared RNA-seq supported introns (red thick lines) and respective number of supported spliced reads are indicated.

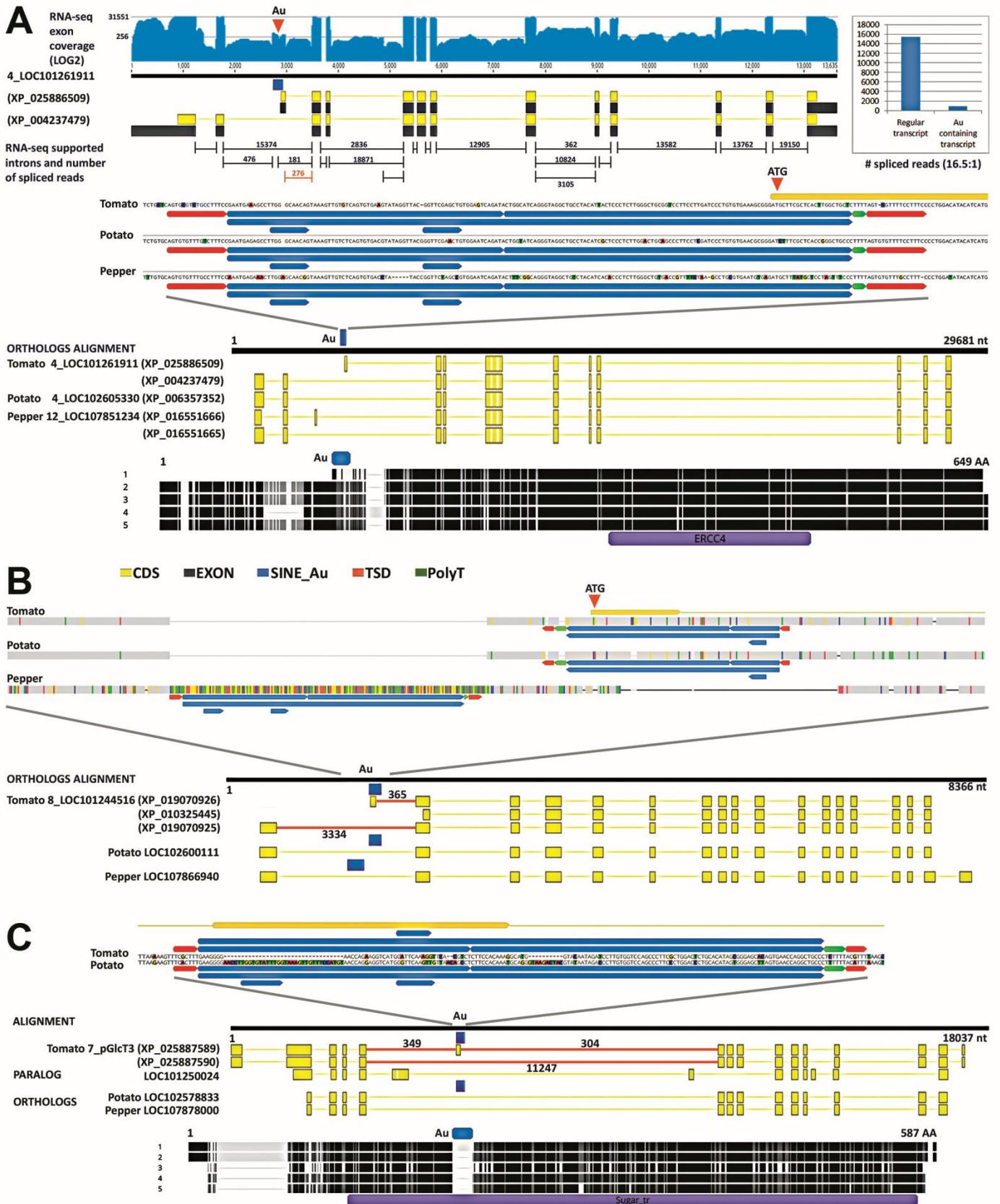


Figure 5. Au SINE annotated tomato gene models showing Au contribution to CDSs and protein variants. A) Alignment of two variants of the crossover junction endonuclease *EME1B* gene with respective orthologs. RNA-seq graphs and data correspond to NCBI Gene Database format. B) Alignment of three variants of the *tpx2* protein gene with respective orthologs. C) Alignment of two variants of the plastidic glucose transporter 2-like protein gene, its paralog and orthologs. Au SINE alignment region is amplified in A-C. Note annotated protein variants of tomato and those from paralog/orthologs in A and C. Compared RNA-seq supported introns (red thick lines) and respective number of supported spliced reads in B and C are indicated.

Table 2. Tomato protein coding genes with SINE_Au at exonic regions. Chr: chromosome. Loci named with numbers are headed by prefix "LOC".

Chr	Gene	Product	Chr	Gene	Product
Au contribution to UTR length and composition					
1	101261548	pentatricopeptide repeat-containing protein at1g26900, mitoch.	6	101256093	pentatricopeptide repeat-containing protein At5g59600
1	101250046	uncharacterized protein 101250046	6	101261037	pentatricopeptide repeat-containing protein at4g21065-like
1	101257414	late blight resistance protein homolog r1a-3	6	101263323	silicon efflux transporter lsi2-like
2	101249596	probable starch synthase 4, chloroplastic/amyloplastic	7	101264838	ethylene-responsive transcription factor erf054
2	109119610	patatin-like protein 2	7	101257484	integrin-linked protein kinase 1
3	101251446	neutral ceramidase 1	8	101255019	cyclin-dependent kinase f-4-like
3	101249541	uncharacterized protein 101249541	8	101255324	uncharacterized protein 101255324
3	101252255	plant-specific tfiib-related protein ptf2	8	101261938	short-chain dehydrogenase tic 32, chloroplastic-like
3	101256255	atp sulfurylase 1, chloroplastic-like	9	101263849	phosphoglucan, water dikinase, chloroplastic
3	101264494	trna-specific adenosine deaminase tad3	9	101266152	wall-associated receptor kinase-like 20
3	101268270	uncharacterized protein 101268270	10	leglo2	glycolate oxidase
4	101258246	microtubule-associated protein 70-5-like	10	sun	calmodulin binding protein sun-like
4	101266113	cd2 antigen cytoplasmic tail-binding protein 2	11	101248259	aminoacylase-1
4	101266811	zinc finger protein 4	11	101252136	protein decreased size exclusion limit 1
4	101257061	peptidyl-prolyl cis-trans isomerase fkbp16-3, chloroplastic	11	101256646	protein suppressor of k(+) transport growth defect 1
5	101248009	cyclic dof factor 2	11	pglcat8	glycosyltransferase
5	101247126	oberon-like protein	12	101249332	protein fip1
5	101246532	transcriptional adapter ada2b	12	101248458	dna-directed rna polymerases i and iii subunit rpc1
5	104647293	transcription factor tcp12-like	12	101254074	uncharacterized protein 101254074
5	112941468	uncharacterized protein 112941468	12	101260319	xyloglucan-specific galacturonosyltransferase 1-like
5	101250507	folylpolyglutamate synthase	12	101247294	nuclear transcription factor y subunit a-7-like
5	101251378	pentatricopeptide repeat-containing protein at1g69350, mitoch.	12	101248088	g2/mitotic-specific cyclin-2-like
5	101253286	vinorine synthase	12	101250477	fimbrin-5
6	101252600	rab3 gtpase-activating protein catalytic subunit			
Au contribution to UTR transcript variants (as novel splice site)					
1	101256630	uncharacterized protein 101256630	6	5pt3	inositol-1,4,5-trisphosphate-5-phosphatase
1	101243920	uncharacterized protein 101243920	9	109121036	homeobox-leucine zipper protein hdg11-like
2	tm-1	tm-1 protein	12	101257050	uncharacterized protein 101257050
2	101252742	protein wvd2-like 7			
Au contribution to CDS boundary definition (as novel start*/stop codons**)					
4	101267112**	ent-kaurene oxidase, chloroplastic	12	101250962**	uncharacterized protein 101250962
7	101267329*	protein phosphatase 2c 55	12	101260016**	receptor-like protein kinase 5
10	101254142**	cellulose synthase-like protein d4			
Au contribution to protein domains and also variants (as novel splice*/start codon**)					
4	101264515*	oligopeptide transporter 4-like	5	112941468	uncharacterized protein 112941468
4	101261911***	crossover junction endonuclease eme1b	7	101266741	uncharacterized protein 101266741
4	101266508	serpin-zx	7	pglct3*	plastidic glucose transporter 2-like
5	104647255	uncharacterized protein 104647255	8	101244516***	protein tpx2

Funding: This research was funded by AGENCIA NACIONAL DE PROMOCIÓN CIENTÍFICA Y TECNOLÓGICA (ANPCyT-Argentina), grant number UNaM PICT 2014-3328 Préstamo BID N° AR-L 1181.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

REFERENCES

1. Wessler SR. Transposable elements and the evolution of eukaryotic genomes. *Proc Natl Acad Sci USA*. 2006;103(47):17600-1.
2. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*. 2002;297(5588):1833-37.
3. Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie WR, et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature*. 2004;430:471-6.
4. Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijó JA, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*. 2009;136(3):461-72.
5. Tenailon MI, Hollister JD, Gaut BS. A triptych of the evolution of plant transposable elements. *Trends Plant Sci*. 2010;15(8):471-8.
6. Rebollo R, Romanish MT, Mager DL. Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet*. 2012;46:21-41.
7. Kramerov DA, Vassetzky NS. Short retroposons in eukaryotic genomes. *Int Rev Cytol*. 2005;247:165-221.
8. Deragon J-M, Zhang X. Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. *Syst Biol*. 2006;55(6):949-56. Sakowicz T, Gadzalski M, Pszczolkowski W. Short interspersed elements (SINEs) in plant genomes. *Adv. Cell Biol*. 2009;1:1-12. doi: 10.2478/v10052-009-0002-x
9. Wenke T, Dobel T, Sorensen TR, Junghans H, Weisshaar B, Schmidt T. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. *Plant Cell*. 2011;23:3117-28.
10. Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, Schulman AH. Cassandra retrotransposons carry independently transcribed 5S RNA. *Proc Natl Acad Sci USA*. 2008;105(15):5833-38.
11. Kojima KK. A new class of SINEs with snRNA gene-derived heads. *Genome Biol Evol*. 2015;7(6):1702-12.
12. Ohshima K. RNA-mediated gene duplication and retroposons: retrogenes, LINEs, SINEs, and sequence specificity. *Int J Evol Biol*. 2013;424726. doi: 10.1155/2013/424726
13. Roy-Engel AM. A tale of an A-tail. The lifeline of a SINE. *Mob Genet Elements*. 2012;2(6):282-6.
14. Kudo S, Fukuda M. Structural organization of glycophorin A and B genes: glycophorin B gene evolved by homologous recombination at Alu repeat sequences. *Proc Natl Acad Sci USA*. 1989;86:4619-23.
15. Rearden A, Magnet A, Kudo S, Fukuda M. Glycophorin B and glycophorin E genes arose from the glycophorin A ancestral gene via two duplications during primate evolution. *J Biol Chem*. 1993;268(3):2260-67.
16. Lee T-F, Gurazada SGR, Zhai J, Li S, Simon SA, Matzke MA, et al. RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats. *Epigenetics*. 2012;7:781-795.
17. Kralovicova J, Patel A, Searle M, Vorechovsky I. The role of short RNA loops in recognition of a single-hairpin exon derived from a mammalian-wide interspersed repeat. *RNA Biol*. 2015;12(1):54-69.
18. Arnaud P, Goubely C, Pélissier T, Deragon J-M. SINE retroposons can be used in vivo as nucleation centers for de novo methylation. *Mol Cell Biol*. 2000;20(10):3434-41.
19. Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D, et al. Minimal conditions for exonization of intronic sequences: 5' splice site formation in Alu exons. *Mol Cell*. 2004;14(2):221-31.
20. Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, et al. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J*. 2006;49(1):38-45.
21. Wick N, Luedemann S, Vietor I, Cotton M, Wildpaner M, Schneider G, et al. Induction of short interspersed nuclear repeat-containing transcripts in epithelial cells upon infection with a chicken adenovirus. *J Mol Biol*. 2003;328(4):779-790.
22. Pouch-Pelissier M-N, Pelissier T, Elmayan T, Vaucheret H, Boko D, Jantsch MF, et al. SINE RNA induces severe developmental defects in *Arabidopsis thaliana* and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet*. 2008;4(6):e1000096. doi: 10.1371/journal.pgen.1000096
23. Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, et al. Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat Commun*. 2014;5:4027. doi: 10.1038/ncomms5027

24. Wikstrom N, Savolainen V, Chase MW. Evolution of the angiosperms: calibrating the family tree. *Proc Biol Sci.* 2001;268(1482):2211-20.
25. Yasui Y, Nasuda S, Matsuoka Y, Kawahara T. The Au family, a novel short interspersed element (SINE) from *Aegilops umbellulata*. *Theor Appl Genet.* 2001;102:463-70.
26. Fawcett JA, Kawahara T, Watanabe H, Yasui Y. A SINE family widely distributed in the plant kingdom and its evolutionary history. *Plant Mol Biol.* 2006;61:505-14.
27. Yagi E, Akita T, Kawahara T. A novel Au SINE sequence found in a gymnosperm. *Genes Genet Syst.* 2011;86(1):19-25.
28. Seibt KM, Wenke T, Muders K, Truberg B, Schmidt T. Short interspersed nuclear elements (SINEs) are abundant in Solanaceae and have a family-specific impact on gene structure and genome organization. *Plant J.* 2016;86:268-85.
29. Keidar D, Doron C, Kashkush K. Genome-wide analysis of a recently active retrotransposon, Au SINE, in wheat: content, distribution within subgenomes and chromosomes, and gene associations. *Plant Cell Rep.* 2018;37:193-208.
30. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5(3):e9490. doi: 10.1371/journal.pone.0009490
31. plants.ensembl.org/ [Internet]. European Molecular Biology Laboratory's European Bioinformatics Institute; c2021 [cited 2021 Jan 09]. Available from: <http://plants.ensembl.org/>
32. pfam.xfam.org/ [Internet]. The protein families database. European Molecular Biology Laboratory's European Bioinformatics Institute; c2020 [cited 2021 Jan 09]. Available from: <http://pfam.xfam.org/>
33. ncbi.nlm.nih.gov/gene/ [Internet]. National Center for Biotechnology Information, US National Library of Medicine; [cited 2021 Jan 09]. Available from: <https://www.ncbi.nlm.nih.gov/gene/>
34. amigo.geneontology.org/amigo [Internet]. The Gene Ontology; c1999-2020 [cited 2021 Jan 09]. Available from: <http://amigo.geneontology.org/amigo>
35. Sánchez DM. Evolución y mapeo de elementos transponibles cortos de la familia AuSINE en tomate. Bachelor in Sciences (Genetics) Dissertation. National University of Misiones, Argentina. 2015;78 pp. Spanish.
36. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature.* 2012;485:635-641.
37. Erdmann RM, Picard CL. RNA-directed DNA methylation. *PLoS Genet.* 2020;16(10):e1009034. doi: 10.1371/journal.pgen.1009034
38. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Pérez JL, Moran JV. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol Spectr.* 2015;3(2):MDNA3-0061-2014. doi: 10.1128/microbiolspec.MDNA3-0061-2014.
39. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. *BioEssays.* 2000;22:148-60.
40. Chang S-B, Yang T-J, Datema E, van Vugt J, Vosman B, Kuipers A, et al. FISH mapping and molecular organization of the major repetitive sequences of tomato. *Chromosome Res.* 2008;16:919-33.
41. Wang X, Ai G, Zhang C, Cui L, Wang J, Li H, et al. Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopodium-specific lncRNAs in tomato. *New Phytol.* 2016; 209(4):1442-55.
42. Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV. Prevalence of intron gain over intron loss in the evolution of paralogous gene families. *Nucleic Acids Res.* 2004;32(12):3724-33.
43. Sugnet CW, Srinivasan K, Clark TA, O'Brien G, Cline MS, Wang H, et al. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput Biol.* 2006;2(1):e4. doi: 10.1371/journal.pcbi.0020004
44. Dumesic PA, Madhani HD. The spliceosome as a transposon sensor. *RNA Biol.* 2013;10(11):1653-60.
45. Hirose T, Mishima Y, Tomari Y. Elements and machinery of non-coding RNAs: toward their taxonomy. *EMBO Rep.* 2014;15:489-507.
46. Jo S-B, Choi SS. Introns: The functional benefits of introns in genomes. *Genomics Inform.* 2015;13:112-118.
47. Rose AB. Introns as gene regulators: A brick on the accelerator. *Front Genet.* 2019;9:672. doi: 10.3389/fgene.2018.00672
48. Rearick D, Prakash A, McSweeney A, Shepard SS, Fedorova L, Fedorov A. Critical association of ncRNA with introns. *Nucleic Acids Res.* 2011;39(6):2357-66.
49. Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci.* 2003;28(4):215-220.
50. Rose AB. Intron-mediated regulation of gene expression. *Curr Top Microbiol Immunol.* 2008;326:277-290.

51. Rigau M, Juan D, Valencia A, Rico D. Intronic CNVs and gene expression variation in human populations. *PLoS Genet.* 2019;15(1):e1007902. doi: 10.1371/journal.pgen.1007902
52. Poverennaya IV, Roytberg MA. Spliceosomal introns: Features, functions, and evolution. *Biochemistry (Mosc).* 2020;85(7):725-34.
53. Zhang Q, Li H, Zhao X-q, Xue H, Zheng Y, Meng H, et al. The evolution mechanism of intron length. *Genomics.* 2016;108(2):47-55.
54. Carvunis A-R, Rolland T, Wapinski I, Carlderwood MA, Yildirim MA, Simonis N, et al. Proto-genes and de novo gene birth. *Nature.* 2012;487(7407):370-4.
55. Comeron JM, Williford A, Kliman RM. The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity.* 2008;100:1931.
56. Monteuiis G, Justin J.L. Wong JLL, Charles G. Bailey CG, Schmitz U, et al. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res.* 2019;47(22):11497-513.
57. Pelissier T, Bousquet-Antonelli C, Lavie L, Deragon J-M. Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res.* 2004;32(13):3957-66.
58. Louro R, Smirnova AS, Verjovski-Almeida S. Long intronic noncoding RNA transcription: Expression noise or expression choice? *Genomics.* 2009;93:291-298.
59. Hesselberth JR. Lives that introns lead after splicing. *WIREs RNA.* 2013;4:677-691.
60. Buckley PT, Khaladkar M, Kim J, Eberwine J. Cytoplasmic intron retention, function, splicing, and the sentinel RNA hypothesis. *WIREs RNA.* 2014;5(2):223-230.
61. Li J, Wang Z, Peng H, Liu Z. A MITE insertion into the 3'-UTR regulates the transcription of TaHSP16.9 in common wheat. *Crop J.* 2014;2(6):381-387.
62. Bogard B, Francastel C, Hubé F. A new method for the identification of thousands of circular RNAs. *Noncoding RNA Investig.* 2018;2:5. doi: 10.21037/ncri.2018.01.02
63. Panda AC, Gorospe M. Identifying intronic circRNAs: progress and challenges. *Noncoding RNA Investig.* 2018;2:34. doi: 10.21037/ncri.2018.05.06
64. Litholdo Jr. CG, Cordenonsi da Fonseca G. Circular RNAs and plant stress responses. In: Xiao J, editor. *Circular RNAs. Advances in Experimental Medicine and Biology*, vol 1087. Singapore: Springer; 2018. p. 345-353.
65. Tan J, Zhou Z, Niu Y, Sun X, Deng Z. Identification and functional characterization of tomato circRNAs derived from genes involved in fruit pigment accumulation. *Sci Rep.* 2017;7:8594. doi: 10.1038/s41598-017-08806-0
66. Wang J, Yang Y, Jin L, Ling X, Liu T, Chen T, et al. Re-analysis of long non-coding RNAs and prediction of circRNAs reveal their novel roles in susceptible tomato following TYLCV infection. *BMC Plant Biol.* 2018;18:104. doi: 10.1186/s12870-018-1332-3
67. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with Alu repeats. *RNA.* 2013;19(2):141-57.
68. Chen L, Zhang P, Fan Y, Lu Q, Li Q, Yan J, et al. Circular RNAs mediated by transposons are associated with transcriptomic and phenotypic variation in maize. *New Phytol.* 2018;217:1292-306.
69. Liu H, Yin J, Xiao M, Gao C, Mason AS, Zhao Z, et al. Characterization and evolution of 5' and 3' untranslated regions in eukaryotes. *Gene.* 2012;507:106-11.
70. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene.* 2001;276(1-2):73-81. doi: 10.1016/s0378-1119(01)00674-6
71. Drongitis D, Aniello F, Fucci L, Donizetti A. Roles of transposable elements in the different layers of gene expression regulation. *Int J Mol Sci.* 2019;20:5755. doi:10.3390/ijms20225755
72. Shen J, Liu J, Xie K, Xing F, Xiong F, Xiao J, et al. Translational repression by a miniature inverted-repeat transposable element in the 3' untranslated region. *Nat Commun.* 2017;8:14651. doi:10.1038/ncomms14651
73. Scarpato M, Angelini C, Cocca E, Pallotta MM, Morescalchi MA, Capriglione T. Short interspersed DNA elements and miRNAs: a novel hidden gene regulation layer in zebrafish? *Chromosome Res.* 2015;23:533-544.
74. Niu X-M, Xu Y-C, Li Z-W, Bian Y-T, Hou X-H, Chen J-F, et al. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc Natl Acad Sci USA.* 2019;116(14):6908-13.
75. Browning JW, Rambo TME, McKay BC. Comparative genomic analysis of the 3' UTR of human MDM2 identifies multiple transposable elements, an RLP24 pseudogene and a cluster of novel repeat sequences that arose during primate evolution. *Gene.* 2020;741:144557. doi: 10.1016/j.gene.2020.144557
76. Maquat LE. Short interspersed nuclear element (SINE)-mediated post-transcriptional effects on human and mouse gene expression: SINE-UP for active duty. *Phil Trans R Soc B.* 2020;375:20190344. doi: 10.1098/rstb.2019.0344
77. Ishibashi K, Ishikawa M. The resistance protein Tm-1 inhibits formation of a Tomato Mosaic Virus replication protein-host membrane protein complex. *J Virol.* 2013;87(14):7933-39.

78. Hughes TA. Regulation of gene expression by alternative untranslated regions. *Trends Genet.* 2006;22(3):119-122.
79. Kim S, Park J, Yeom S-I, Kim Y-M, Seo E, Kim K-T, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 2017;18:210. doi: 10.1186/s13059-017-1341-9
80. Pray L. Functions and utility of Alu jumping genes. *Nature Education.* 2008;1(1):93.
81. Schmitz J, Brosius J. Exonization of transposed elements: A challenge and opportunity for evolution. *Biochimie.* 2011;93(11):1928-34.



© 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).