

Article

A Global Landscape of Miniature Inverted-Repeat Transposable Elements in the Carrot Genome

Alicja Macko-Podgórní *^{ID}, Gabriela Machaj and Dariusz Grzebelus ^{ID}

Department of Plant Biology and Biotechnology, University of Agriculture in Krakow, 31-120 Kraków, Poland; g.machaj@student.urk.edu.pl (G.M.); d.grzebelus@urk.edu.pl (D.G.)

* Correspondence: a.macko@urk.edu.pl; Tel.: +48-12-662-51-89

Abstract: Miniature inverted-repeat transposable elements (MITEs) are the most abundant group of Class II mobile elements in plant genomes. Their presence in genic regions may alter gene structure and expression, providing a new source of functional diversity. Owing to their small size and lack of coding capacity, the identification of MITEs has been demanding. However, the increasing availability of reference genomes and bioinformatic tools provides better means for the genome-wide identification and analysis of MITEs and for the elucidation of their contribution to the evolution of plant genomes. We mined MITEs in the carrot reference genome DH1 using MITE-hunter and developed a curated carrot MITE repository comprising 428 families. Of the 31,025 MITE copies spanning 10.34 Mbp of the carrot genome, 54% were positioned in genic regions. *Stowaways* and *Tourists* were frequently present in the vicinity of genes, while *Mutator*-like MITEs were relatively more enriched in introns. *hAT*-like MITEs were relatively more frequently associated with transcribed regions, including untranslated regions (UTRs). Some carrot MITE families were shared with other Apiaceae species. We showed that *hAT*-like MITEs were involved in the formation of new splice variants of insertion-harboring genes. Thus, carrot MITEs contributed to the accretion of new diversity by altering transcripts and possibly affecting the regulation of many genes.

Keywords: *Daucus carota*; genome annotation; MITE; alternative splicing; gene expression



Citation: Macko-Podgórní, A.; Machaj, G.; Grzebelus, D. A Global Landscape of Miniature Inverted-Repeat Transposable Elements in the Carrot Genome. *Genes* **2021**, *12*, 859. <https://doi.org/10.3390/genes12060859>

Academic Editor: Ingo Schubert

Received: 10 May 2021

Accepted: 2 June 2021

Published: 3 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Miniature inverted-repeat transposable elements (MITEs) are non-autonomous Class II mobile elements, widespread and abundant in plant genomes. MITEs constitute an artificial group within class II elements characterized by high copy numbers, small size (>700 bp) and lack of any coding capacity [1]. They are further divided into sub-groups based on the similarity of their terminal inverted repeats (TIRs) and target site duplications (TSDs) to those of certain elements representing major superfamilies of Class II transposons. *Tourists* and *Stowaways* were the first described and the most thoroughly characterized MITE sub-groups derived from *PIF/Harbinger* and *Tc1/Mariner* elements, respectively [2,3]. Subsequently, MITEs related to other DNA transposons, i.e., *Mutators* and *hATs*, were also reported [4,5]. The similarity of MITE TIRs to those of autonomous DNA transposons allows MITEs to be mobilized by transposases encoded by their autonomous relatives, resulting in the transposition and formation of characteristic TSDs upon insertion.

MITEs are often inserted in the vicinity of genes. Such a location entails the possibility of a functional impact on the nearby gene. Besides the knockout mutations resulting from a disruption of the gene structure [6–8], numerous reports have been published indicating the possibility that MITE insertions may alter the expression of adjacent genes [9–14]. MITEs can also give rise to siRNAs, and drive gene silencing through RNA-directed methylation (RdDM) [14,15]. It has been shown that genes containing MITE insertions in their promoters may be epigenetically down-regulated, e.g., MITE-associated methylation of *ZmCCT* and *ZmNAC111* promoters in maize resulted in an early-flowering phenotype and a drought-sensitive phenotype of maize seedlings, respectively [16,17]. MITEs may also

carry transcription factor binding sites (TFBS) [18,19] or cis-acting DNA regulatory elements, e.g., new splice sites [20], transcription start sites, TATA boxes, and polyadenylation signals [21]. Insertions of MITEs into 3' UTRs may considerably affect post-transcriptional gene regulation, through modulation of RNA stability or translation [21–23], while insertion into the upstream region of a gene may result in its up-regulation [24,25]. However, in order to investigate the impact of MITEs on host genes, their genome-wide identification and characterization is an essential prerequisite.

Plant genomes differ in terms of MITE content and diversity. The number of MITE copies can vary sharply even among closely related species. For example, the number of MITE copies in the genome of *Arabidopsis lyrata* is more than five times higher than in *Arabidopsis thaliana*, while the genome of the former is only twice as large as that of the latter [26]. Similarly, watermelon, having a slightly smaller genome than melon, contains seven times more MITE copies [26]. High levels of MITE diversity have been observed in Poaceae, while within dicot plants, species characterized by high MITE diversity are less common [26]. For example, in the grapevine genome, only eight MITE families were identified, and neither *Stowaway* MITEs nor related autonomous *Tc1/Mariners* were found [27,28]. It is not clear what mechanism is responsible for such diversification among plants.

The carrot (*Daucus carota* subsp. *sativus* Hoffm.) is an economically important crop, and a significant source of β -carotene in the human diet [29]. Although the carrot has a relatively small genome of ca. 473 Mb, the repetitive fraction constitutes around 46%, of which 13.6% is attributed to Class II DNA transposons [30]. To date, two groups of carrot MITEs have been more thoroughly characterized, i.e., *Krak* elements belonging to the *Tourist* group [31] and *DcSto* elements attributed to the *Stowaway* group [32,33]. Owing to their high insertional polymorphism, they were utilized as sources of molecular markers [34–36]. It was shown that *DcSto* elements were associated with genes, most frequently occurring in 5' and 3' UTRs. A significant enrichment of gene encoding transcription factors was also revealed in the fraction of genes associated with *DcSto* insertions in their 2 kb up- and downstream regions, suggesting the possible functional impact of those MITEs [33]. Nevertheless, a comprehensive, global annotation of carrot MITEs has been lacking.

Here, we report on the global characterization of carrot MITEs, their genomic distribution, association with genes and presence in transcripts. We provided examples of genic insertions of *hAT*-like MITEs resulting in the formation of novel splice variants. We also identified MITEs in genomes of other Asterid species and showed that the carrot, despite its relatively small genome, was characterized by an exceptional diversity and abundance of MITEs. The curated carrot MITE repository will facilitate research on the origin of carrot MITEs and their role in the host genome.

2. Materials and Methods

2.1. Identification and Genomic Localization of MITEs in the Carrot Reference Genome

Carrot MITE families were identified with MITE-Hunter [37], using default parameters. The final dataset, comprising 522 MITE consensus sequences, was manually curated and grouped into 428 MITE families using the 80-80-80 similarity rule [1]. Those sequences were used as queries for a *blastn* search [38] against the carrot DH1 reference genome (GenBank assembly accession number: GCA_001625215.1; [30]). The *blastn* output was parsed using a custom Perl script, with parameters allowing the extraction of coordinates of copies following 80–80–80 similarity rule, and for which similarity at both ends started between 1–10 nucleotides and continued over the whole query sequence. Genomic regions meeting those criteria were used to produce two bed files, one with exact coordinates and the other with coordinates extended by 50 nt at both ends. The bed file with the MITE-flanking regions was used to extract sequences that were manually inspected to verify/identify MITE TIRs and TSDs. Subsequently, TSDs were used to combine families into groups related to DNA transposons [Class II], i.e., *Stowaways* (related to *Tc1/mariner*), *Tourists* (related to *PIF/Harbinger*), *hAT*-like, and *Mutator*-like MITEs.

To mine for autonomous elements related to MiMs (MITEs inserted in microsatellite) [26], we used *blastx* implemented in *blastall* (v 1 -b 1) to search for *Mutator* transposases in the carrot reference genome. As queries, we used *Mutator* proteins identified in *Oryza sativa* (AAX92869.1), *Rosa chinensis* (PRQ51974.1), *Helianthus annuus* (OTF91787.1), *Solanum tuberosum* (ABI34394.1), *Arabidopsis thaliana* (AAG51216.1), and *Medicago truncatula* [39]. Genomic regions with similarity to the *Mutator* proteins were extracted with 4 kb flanking sequences and used to identify TIRs. TIRs were recognized based on the results of the self-blast of extracted sequences (*blastn*; -v 1 -b 1), from which the coordinates of hits present on both flanks of the putative transposase coding region were in complementary orientation and shared at least 50% identity. They were used to define the boundaries of putative autonomous elements. Subsequently, we used *blastn* with default parameters to compare 50-bp-long ends of consensus MiM sequences with the mined *Mutator*-like elements.

A bed file containing coordinates of MITEs was used to determine the positions of MITE copies in the context of genic regions, as described by Macko-Podgórni et al. [33]. MITE copies were categorized based on the NCBI carrot genome annotation (GCF_001625215.1_ASM162521v1_genomic.gff) as 5'UTR (MITE coordinates overlapping annotated 5'UTRs), 3'UTR (MITE coordinates overlapping annotated 3'UTRs), cds (MITE coordinates overlapping annotated cds), intron (MITE coordinates located within an annotated intron), upstream (MITE present within 1 kb upstream a gene), downstream (MITE present within 1 kb downstream a gene), while copies not fulfilling any of those criteria were considered as intergenic. All calculations were performed using the bed file in R [40]. A gff3 file showing the annotation of carrot MITEs in the reference genome was developed (File S1 in the Supplementary Materials).

2.2. Carrot MITEs in Transcribed Regions

RNAseq reads from twenty DH1 tissues (PRJNA291977; Supplementary Table S2, [30]) were used to identify MITE-containing transcripts. Reads mapping to MITE sequences were found using *Bowtie2* v. 2.3.5.1 [41], extracted with *SAMtools* v. 1.9 [42], blasted against the carrot DH1 reference genome with *blastall* [38] with the following parameters: -p *blastn* -F F -v 1 -b 1 -K 1. Blast hits longer than 20 bp with a similarity higher than 80% were kept. Subsequently, regions with blast hits were overlapped with MITE annotation using *BedTools* v. 2.26.0 [43]. If at least ten reads mapped to the MITE sequence, the corresponding gene with a MITE insertion (upstream, downstream, in 5'UTR, 3'UTR, cds, or intron) was further analyzed.

To determine the expression of MITE-containing and non-MITE-containing isoforms, RNAseq reads from DH1 were mapped to the carrot DH1 reference genome (GenBank assembly accession number: GCA_001625215.1 [30], Table S1 in the Supplementary Materials) using *STAR* v. 2.7.3a [44] with the following parameters: *outSAMmapqUnique*: 50; *outFilterMultimapNmax*: 20; *alignSJoverhangMin*: 8; *alignSJDBoverhangMin*: 1; *outFilterMismatchNmax*: 999; *outFilterMismatchNoverLmax*: 0.04; *alignIntronMin*: 20; *alignIntronMax*: 1,000,000; and *alignMatesGapMax*: 1,000,000. Novel isoform identification and calculation were performed according to Machaj et al. [45]. Subsequently, *SAMtools* v. 1.9 [42] and *BedTools* v. 2.26.0 [43] were used to extract MITE and gene coverage from bam files. MITE/gene (M/g) ratios were calculated based on normalized RPK (reads per kilobase; RPK/(library size/1000000)) values and genes with RPK > 1 and the M/g ratio > 1.9 were further evaluated. Analysis of the differential exon usage, resulting from the presence of a MITE copy, was performed using *JunctionSeq* [46] based on RNAseq data from seven carrot tissues differing with respect to the number of normalized RNAseq reads attributed to MITEs. The differential exon usage was tested between “bracts (from not opened flower), from 2 cm umbel” (SRR2148980) as a reference, and each of the six tissues: “callus” (SRR2148992), “fibrous roots” (SRR2148991), “whole flowers (not opened), 2 cm umbel” (SRR2148981), “leaves stage 1, 0.5–1 cm young sprout” (SRR2148984), “stressed root, whole storage root” (SRR2148997), and “germinating seeds, at the beginning of germination” (SRR2148999). In total, 40 genes carrying copies of *hAT*-like MITEs were

investigated. Expression levels of those genes were represented as a heatmap, created in R [40].

2.3. Mining for MITEs in the Reference Genomes of Other Asterid Species

The mining strategy used to identify carrot MITEs was applied to eight reference genomes representing members of the Asterid clade (Supplementary Table S2 [47–53]). Then, *blastn* [38] was used for among-species comparisons of MITE consensus sequences to identify MITE families shared among Asterids. The presence of a region covering a minimum of 60% MITE length with identity equal to or higher than 60% was used as a threshold to define related MITE families from different genomes. A graphical representation of levels of similarity among related MITEs was drawn using *Circoletto* [54], with the following *blastn* parameters: *-F F -e 1e-10 -E -1 -v 200 -b 200*, and “score/max” ratio coloring with blue ≤ 0.25 , green ≤ 0.50 , orange ≤ 0.75 , red > 0.75 .

3. Results

3.1. Abundance and Genomic Localization of Carrot MITEs

We identified 428 MITE families in the carrot reference genome and divided them into groups on the basis of their relationships to Class II TE superfamilies, as revealed by their TIR and TSD similarity (Figure 1). We used consensus sequences representing each family to identify individual copies along the reference assembly of the carrot genome [30]. In total, MITEs were estimated to occupy around 2% of the carrot genome (31,025 copies spanning 10.34 Mbp).

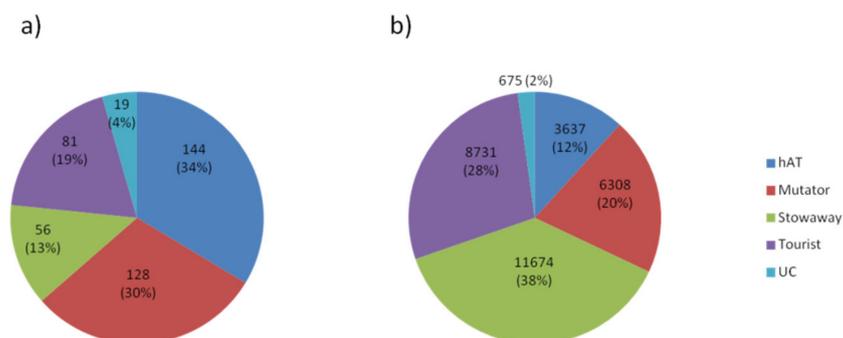


Figure 1. Abundance of MITE groups in the carrot genome with respect to the number of families (a) and the number of copies (b). UC stands for unclassified MITEs.

The largest genome fraction was attributed to *Stowaways* (2.97 Mbp) followed by *Tourists* (2.78 Mbp) and *Mutator*-like MITEs (2.71 Mbp). *hAT*-like MITEs were the least numerous but the most diverse, comprising one-third of all identified MITE families. Families derived from *Mutator*-like elements accounted for ca. 30%, while *Tourists* and *Stowaways* grouped 19% and 13% of MITE families, respectively (Table 1). More than half of all MITEs (16,693 copies; 54%) were inserted within genic regions, defined as 1 kb upstream and downstream of genes and including the gene body (Table 1). We also identified 2156 copies belonging to 15 families flanked by “TA” stretches and having distinguishable but usually poorly conserved TIRs. For three of those families, we were able to identify related putative autonomous *Mutator* elements, confirming that they should be attributed to the group of *Mutator*-like MITEs, and classified them as MiMs, following the nomenclature of Chen et al. [26].

Table 1. MITEs identified in the carrot genome.

MITE	Number of Families	Number of Copies	Mbp	% of Genome	Mean Copy Number per Family	Mean Length of MITE (bp)	Copies in Most Numerous Family	Genomic Localization						
								Intergenic	Upstream	5'UTR	Cds	Intron	3'UTR	Downstream
<i>hAT</i>	144	3637	1.65	0.35%	25	454	217	1506	608	77	31	918	66	431
<i>Mutator</i>	128	6308	2.71	0.57%	49	429	542	3019	1148	52	12	1343	58	676
<i>Stowaway</i>	56	11,674	2.97	0.63%	208	254	1452	5605	2446	109	11	1991	89	1423
<i>Tourist</i>	81	8731	2.78	0.59%	108	319	1353	3867	1803	134	19	1574	110	1224
Unclassified	19	675	0.23	0.05%	36	335	141	335	99	5	2	188	5	41
Total	428	31,025	10.34	2.19%	-	-	-	14,332	6104	377	75	6014	328	3795

The MITE groups differed in terms of their copy number, the average length of the element and genomic localization. *Stowaways* were on average 254 bp-long, and constituted the most abundant group of MITEs in the carrot genome, reaching 11,674 copies. The largest family, comprising 1452 copies, was attributed to that group. The most numerous *Stowaway* families included previously described *DcSto* families [33]. *Stowaway* elements were relatively enriched in intergenic regions and upstream from genes. *Tourists* were represented by 8731 copies with an average element length of 319 bp. The largest family in that group comprised 1353 copies. A previously described *Krak* family [31], comprising 333 copies in the carrot reference genome, was the seventh-largest family of *Tourists*. *Tourists* were more frequently positioned upstream and downstream from genes and in UTRs. *Mutator*-like MITEs comprised 6308 copies, with a mean length of 429 bp. The largest family within that group contained 542 elements. *Mutator*-like MITEs, including MiMs, were present mainly in intergenic regions and in introns. The group of *hAT*-like elements was the least numerous, represented by 3637 copies with an average length of 454 bp and the largest family comprising 217 copies. *hAT*-like elements were relatively more frequently located in coding regions and UTRs, as compared to the other MITE groups (Table 1, Figure 2, Figure S1, File S2 in the Supplementary Materials).

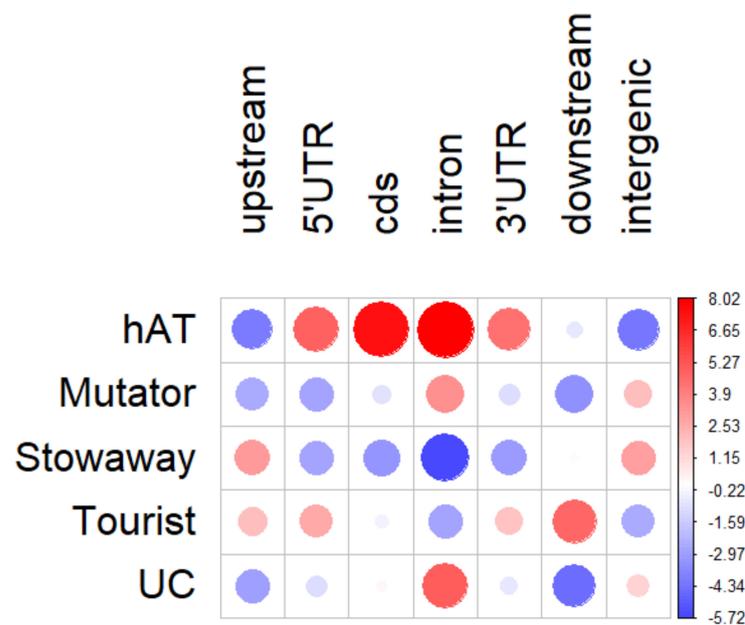


Figure 2. Relative abundance of MITE groups in different genomic regions (p -value = 0.0005). The color scale reflects deviations from average values. Circle size is proportional to the contribution of each test to the total Pearson chi-squared score. UC stands for unclassified MITEs.

3.2. Carrot MITEs Are Co-Transcribed with Genes and Are Involved in Tissue-Specific Alternative Splicing

In order to determine whether MITEs localized in the vicinity of genes were co-transcribed, we used RNAseq reads from 20 carrot tissues [30] and mapped them onto MITE sequences to search for MITEs covered by more than ten reads. In total, 3469 MITE copies, representing 336 families, were retrieved. Of those, 60% were localized in introns, 16% and 14% 1 Kb upstream or downstream from genes, respectively, 4% in 5'UTR, 5% in 3'UTR, and 1% in cds (Table S3, Figure S2 in the Supplementary Materials). The number of copies belonging to particular MITE families, residing in transcribed regions, correlated with the total number of copies in those families (p -value = 2.2×10^{-16} (Supplementary Figure S3)). *hAT*-like elements, which were shown to be relatively more frequent in gene bodies, were also more frequently co-transcribed. Interestingly, *hAT*-like elements located in introns were frequently retained in transcripts, while intronic *Stowaways* and

Mutator-like MITEs were usually spliced out together with introns and did not significantly contribute to transcripts (Figure S4 in the Supplementary Materials). Thus, we were prompted to investigate if *hAT*-like MITEs might be involved in the formation of novel gene-splicing variants.

In some tissues, we observed more reads mapping to MITEs, possibly indicating a higher expression of MITE-containing isoforms. The highest normalized RPK values were shown for “stressed leaves of 7–8 cm at reversible wilting point”, “stressed leaves of 2–2.5 cm at reversible wilting point” and “fibrous roots” (Figure S5 in the Supplementary Materials). Because such differences might have reflected general differences in gene expression, not related to the presence of MITEs, we calculated the MITE vs. gene (M/g) expression ratio, in which a normalized number of reads mapping to the co-transcribed MITE segment was divided by a normalized number of reads mapping to the whole gene, providing means to determine if the MITE-derived isoform was indeed differentially expressed in a tissue-specific manner. In total, we identified 3022 genes with normalized RPK > 1 and M/g > 1.9. Of those, 63%, 26% and 11% of genes carried MITEs localized in introns, upstream or downstream regions of genes, and transcripts, respectively (Table S4, File S2 in the Supplementary Materials).

To test the possible impact of MITEs on the differential usage of exons in different tissues, we retrieved genes for which differences in the maximum and minimum M/g values were higher than 1.5. From those, we randomly selected 40 genes carrying *hAT*-like elements in their UTRs, cds or introns (Table S5 in the Supplementary Materials). We found five genes for which the differential exon usage was reported, for isoforms containing *hAT*-like MITEs (Figure S6–S10; File S3 in the Supplementary Materials). In the case of two genes (LOC108223373 and LOC108201852), the presence of *hAT*-like MITEs led to the formation of isoforms encompassing the whole MITE. The *hAT*-like MITE present in LOC108227539 was partially incorporated into the new exon, while MITEs within the other two genes (LOC108220652 and LOC108222651) were co-transcribed as alternative 5'UTR and 3'UTR, respectively. In general, regardless of the tissue, we observed the prevalence of one isoform. In the case of LOC108222651 and LOC108227539, the non-MITE-containing variant was generally more abundant; however, the expression of MITE-containing isoforms differed slightly among tissues. The expression of LOC108223373 was lower than the other genes (Figure 3a; File S3, in the Supplementary Materials), mostly the MITE-containing isoform was produced, and its expression levels differed depending on the tissue (Figure 3b; File S3 in the Supplementary Materials). For the remaining two genes (LOC108201852 and LOC108220652), MITE-containing isoforms were predominant. The non-MITE-containing isoform of LOC108201852 was expressed only in “bracts (from not opened flower)”, while the MITE-containing isoform was highly expressed in bracts (not open), open flowers, fibrous roots and seeds. The MITE-containing isoform of the LOC108220652 gene was expressed in all tissues, and the highest expression was reported for bracts (not open), open flowers, leaves, hypocotyls, phloem, and xylem. The non-MITE-containing isoform was observed in the callus and leaves.

3.3. Identification of MITEs in Asterids

In order to compare the abundance of MITEs in the carrot genome to those in other Asterid species, we mined MITEs from eight Asterid reference genomes. The number of MITE families and the genome fraction occupied by MITEs were not correlated with the genome size (p -values = 0.37 and 0.79, respectively). The highest abundance and diversity of MITEs (428 families) was observed in the carrot genome, while only 83 MITE families were identified in the celery genome (Figure 4; Table 2).

In addition, the largest proportion of the genome was attributed to MITE elements in the carrot (2.19%, 10.34 Mbp), followed by the sunflower (2.09%, 73.28 Mbp), celery (1.25%, 27.71 Mbp), coffee (0.93%, 12.11 Mbp), potato (0.91%, 7.65 Mbp), and tomato (0.79%, 7.07 Mbp). MITEs were the least abundant in fennel (0.42%, 5.67 Mbp) and pepper (0.36%, 12.43 Mbp). In some investigated genomes, e.g., celery, high MITE copy numbers were

observed despite a limited diversity at the family level (Figure S11 in the Supplementary Materials).

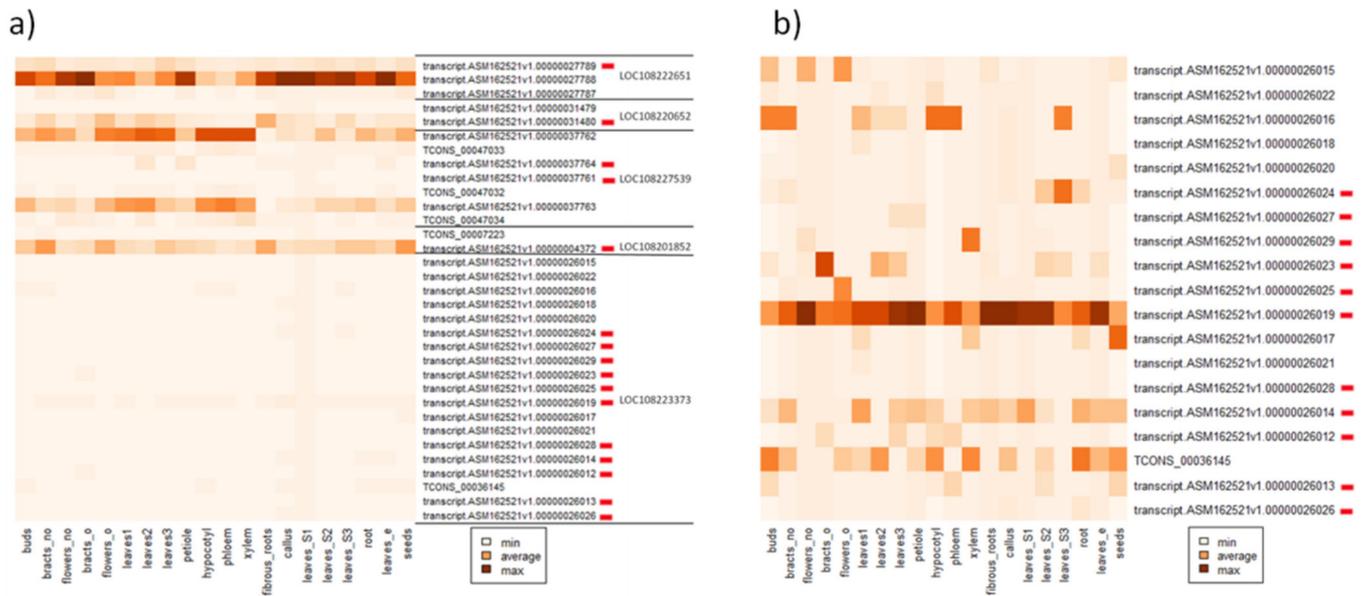


Figure 3. Heatmap of the expression (TPM, transcript per million) of five genes (a) and LOC10822373 (b) producing *hAT*-containing isoforms (labeled by red rectangles). Expression was normalized by tissue type to allow comparisons between genes and isoforms. Novel isoforms are labeled as TCONS_XXX. Detailed results/descriptions are provided in Supplementary File S3.

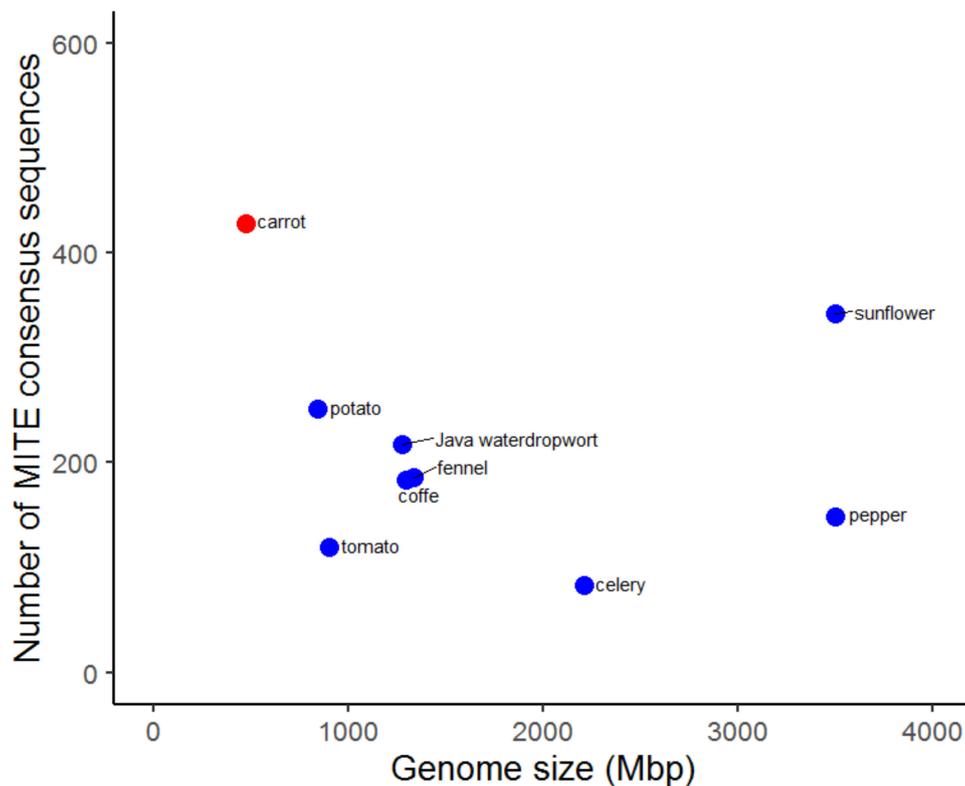


Figure 4. Plot showing the number of MITE families found in Asterid species (*y* axis) related to the genome size (*x* axis).

Table 2. MITEs identified in nine Asterid species.

Plant	Number of MITE Copies	Number of Families (MITE-Hunter)	Occupied Genome Fraction (Mbp)	Fraction of Assembled Genome (%)	Fraction of Genome (%)	Size of Assembly (Mbp)	Genome Size (Mbp)	Reference for the Published Genome Size
celery	52,269	83	27.71	0.83	0.80	3332.58	3470	[47]
pepper	31,405	149	12.43	0.42	0.36	2935.88	3500	[48]
coffee	25,038	184	12.11	1.11	0.93	1094.45	1300	[49]
carrot	31,025	428	10.34	2.45	2.19	421.54	473	[30]
fennel	17,861	186	5.67	0.56	0.42	1010.97	1340	[50]
common sunflower	114,089	342	73.28	2.43	2.09	3010.05	3500	[51]
Java water-dropwort	21,330	217	7.79	0.61	-	1278.51	-	-
tomato	15,150	119	7.07	1.00	0.79	705.93	900	[52]
potato	20,368	251	7.65	0.92	0.91	828.35	844	[53]

MITEs present in genomes of Asterids from different families (Apiaceae, Solanaceae, Rubiaceae, Asteraceae) revealed no similarity above the threshold of 60% sequence identity over 60% of the sequence length. More similar MITEs were identified within Solanaceae (tomato, potato and pepper) and within Apiaceae (carrot, fennel, celery, and Java water dropwort) (Table S6 in the Supplementary Materials). Thus, the number of related MITEs in different genomes corresponded to phylogenetic relationships among the species. In total, similar counterparts in Apiaceae were found for 150 carrot MITE families (Table S7, File S4 in the Supplementary Materials). Of those, nine MITE families were present in the four Apiaceae species, with six families representing *Stowaways* (*DcSto1*, *DcSto4*, *DcSto6*, *DcSto10*, *DcSto29*, and *DcSto37*), two families representing *Tourists*, and one *hAT*-like MITE family (Figure S12 in the Supplementary Materials). Forty-three carrot MITE families were also present in two other Apiaceae species (20 shared with Java water dropwort and fennel, 18 with celery and fennel, and five with Java water dropwort and celery). The remaining 95 carrot MITE families were similar to MITEs from one other Apiaceae species (53, 27 and 15 were shared with fennel, Java water dropwort and celery, respectively).

4. Discussion

4.1. Diversity and Abundance of Carrot MITEs

MITEs are the most abundant group of DNA transposons in plant genomes. Higher diversity and abundance of MITEs have also been reported in monocots than in dicots [26]. We identified 428 families comprising 31,025 copies in the carrot genome. Thus, the carrot stands out from most other dicot species and can be placed among few species, having relatively small genomes but containing large numbers of diverse MITEs, along with *Medicago truncatula* and mulberry as the most prominent examples. In the *M. truncatula* genome of 307 Mbp, 288 MITE families and 132,834 copies were identified [26] while the 357 Mb-long mulberry genome was reported to contain 90,789 full-length copies divided into 232 families [20]. Thus, even though the MITE copy number in the carrot is lower, they show remarkable diversity.

We identified 15 families, grouping more than 2000 MiMs residing in “TA” stretches. MITEs preferably inserted into microsatellites have been described in other plant genomes. Chen et al. [26] reported that MiMs were present in ten of 41 plant species [26]. MiMs were first described in rice as high copy-number MITEs having poorly recognizable TIRs and preferentially inserting into “TA” repeats [55]. MiMs from other species were characterized by similar features and, based on the similarity of their TIRs to related autonomous elements, they were attributed to the *Mutator* superfamily [39,56]. It has been suggested that the propensity to insert into “TA” stretches might reflect a strategy adopted to avoid removal from the genome by positioning into regions not subject to strong selective pressure [39]. However, it remains unclear how those elements propagate in the genome,

whether autonomous *Mutator* elements drive their mobilization, and how mobilized copies are directed to “TA” microsatellites. An alternative mechanism assuming the propagation of MiMs via homologous recombination, preceded by the formation of extrachromosomal circular DNA (eccDNA) intermediates with microsatellite sequences at MiM ends, was proposed by Franco et al. [56].

4.2. Distribution of MITEs in the Carrot Genome

We showed that 16,693 carrot MITEs (54%) reside in genic regions, and 21% of those are co-transcribed with the nearby gene, possibly affecting gene expression or the fate of MITE-containing transcripts. MITEs can interact with genes in multiple ways. Upon insertion in regulatory regions, they may alter gene expression by providing transcription factor binding sites [19]. The presence of polyadenylation signals or splicing sites in the MITE sequence may lead to the formation of new isoforms. On the other hand, MITEs may down-regulate genes by altering methylation through RNA-directed methylation (RdDM) [14,15]. We observed that different carrot MITE groups were enriched in different genic contexts, which may suggest their possible involvement in different regulatory mechanisms. In general, elements inserted into the gene body (UTRs, cds or introns) were often co-transcribed. It was the most apparent in the case of *hAT*-like MITEs, which were inserted into genes relatively more frequently. Here, we demonstrated their engagement in the formation of new isoforms. In the case of two genes, MITE-containing isoforms were prevalent. Such a variability of transcripts may result in a measurable phenotypic effect. For instance, the insertion of a copia retrotransposon into the intron of a tomato gene *Solyc02g079490* produced multiple isoforms without affecting the overall gene expression, but leading to overexpression of non-functional transcripts, ultimately resulting in the accumulation of 2-phenylethanol and giving the tomato fruit a pleasant floral aroma [57]. Carrot *Tourist* elements were preferentially localized in UTRs and regions upstream or downstream from genes, but only insertions residing in UTRs were present in transcripts. In contrast, *Mutator*-like MITEs, enriched not only in intergenic regions but also in introns, were rarely retained in transcripts. *Stowaways* were relatively rarely present in transcripts, despite being the most numerous group of carrot MITEs. However, it should be stressed that the overview of general trends observed for each MITE group was certainly biased by the behavior of the most numerous families. A recent report on carrot *DcStos* indicated strong family-specific preferences for their genomic positioning. Some *DcSto* families were shown to be enriched in genic regions, especially in the vicinity of genes encoding transcription factors [33]. Thus, a detailed characterization of individual families, including information about the expression of nearby genes, is necessary to understand and resolve the involvement of MITEs in the regulation of gene expression.

4.3. MITEs in Other Asterid Species

An *in silico* identification of transposable elements can be biased, and the number of identified repetitive sequences largely depends on the pipeline and stringency of parameters used for analysis. This was why we tested eight Asterid species using the same pipeline we applied for the carrot. The number of families mined from the tomato and potato genomes proved to be consistent with previous reports [26], confirming that the method was adequate, and the carrot genome was indeed characterized by a large diversity and abundance of MITEs. We showed that MITEs were species-specific; however, some MITE families were shared among species from the same botanical family. The number of MITE families shared among species reflected their phylogenetic relationships. We identified 150 carrot MITE families sharing a similarity with MITEs from at least one other Apiaceae species. The intra-family identity in sequence and length decreases over time due to random mutations. The approach we used was based on the 80-80-80 rule for grouping copies into families within the genome, while a less stringent criterion, requiring an identity of at least 60% over 60% of the sequence length, was applied for interspecific comparisons. Detection of nine MITE families shared among Apiaceae, which diverged by

about 26 Mya [58], raises a question about their origin. It is possible that they have been retained owing to their functional role in the host genomes. The TE conservation may be related to wiring new transcriptional networks. For example, similar TFBS-carrying MITEs were reported to be shared among species within the *Prunus* and *Solanum* genera [19]. Similarly, *Helitrons* in *A. thaliana* are enriched in PHE1 binding sites, and their insertions up-regulate a number of genes in the endosperm. Some of those *Helitron* insertions are also present in orthologous genes of other Brassicaceae [59]. It was also shown that some ancient TEs identified in Brassicaceae were enriched in gene regulatory networks, e.g., the flowering gene network, suggesting their domestication and conservation for more than 100 million years [60]. Even though the preservation of MITEs in related species is possible, a horizontal transfer cannot be ruled out, as there has been evidence of TE horizontal transfers in animals and plants, and between viruses and their hosts [61–64].

Our results suggest that some MITEs may play regulatory functions in the carrot and other Apiaceae. We reported a carrot MITE repository and genome annotation providing means for further analysis of MITEs and their effect on the structure and function of carrot genes and genomes. Besides the analysis of the functional role and impact on the genome evolution, insertional polymorphism of MITEs belonging to newly described families can be utilized as molecular markers. Until now, MITE-based molecular markers were successfully used to study carrot diversity and population structure [35,36]. Identification of new MITE families, especially those associated with genes, provides a rich source of transposable element-associated structural variants (TEASVs) that can be used for TE-based genome-wide association mapping (TE-GWAS), an approach recently successfully applied in rice [65].

5. Conclusions

Carrot MITEs are exceptionally diverse and abundant. Their localization in the vicinity of genes and presence in transcribed regions points to their possible involvement in the regulation of gene expression and the formation of novel isoforms. The identification and characterization of carrot MITEs provide a basis for further studies on their functional impact. It also facilitates the use of MITE insertion polymorphisms to identify genetic factors associated with important agronomic traits of the carrot.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12060859/s1>, Figure S1: Box-plot showing the number of copies in each group of carrot MITEs. Figure S2: Abundance and genomic localization of carrot MITEs present in transcripts. Figure S3: Relationship between the number of copies in transcribed regions and the total number of copies representing each MITE family. Figure S4: Relative abundance of MITE groups in different genomic regions, divided into present (bolded) or absent in transcripts (p -value < 0.001). Figure S5: Number of normalized RNAseq reads (RPM) attributed to each MITE group. Figure S6: Splice variants and differential exon usage for LOC108223373. Figure S7: Splice variants and differential exon usage for LOC108201852. Figure S8: Splice variants and differential exon usage for LOC108227539. Figure S9: Splice variants and differential exon usage for LOC108220652. Figure S10: Splice variants and differential exon usage for LOC108222651. Figure S11: Plot showing the number of MITE copies present in Asterid species (y axis) related to the genome size (x axis). Figure S12: Relationships among MITE families shared by four Apiaceae species. Table S1: Carrot DH1 transcriptomes used for identification of MITEs in transcripts. Table S2: Asterid species used for identification of MITEs. Table S3: Carrot MITEs identified in transcripts. Table S4: Summary of MITEs present in transcripts and inserted in the vicinity/within genes with RPK > 1 and M/g > 1.9. Table S5: List of randomly selected genes containing *hAT*-like MITE insertions assigned to UTRs, cds or introns, tested for MITE-related differential exon usage. Table S6: Number of MITE families shared between Asterid species. Table S7: Number of carrot MITE families shared by two or more Apiaceae species. File S1: A gff3 file showing annotation of mined MITEs in the carrot reference genome. File S2: Characteristics of carrot MITE families, File S3: Detailed results of analysis of isoforms of *hAT*-like MITE containing genes. File S4: List of MITE families shared by two or more Apiaceae species.

Author Contributions: Conceptualization, A.M.-P. and D.G.; methodology, A.M.-P.; software, A.M.-P. and G.M.; formal analysis, A.M.-P.; investigation, A.M.-P. and G.M.; resources, A.M.-P.; data curation, A.M.-P.; writing—original draft preparation, A.M.-P.; writing—review and editing, A.M.-P. and D.G.; visualization, A.M.-P. and G.M.; supervision, A.M.-P.; project administration, A.M.-P.; funding acquisition, A.M.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Polish National Science Center (NCN), 2019/33/B/NZ9/00757 (OPUS17), and funds for research at the University of Agriculture in Krakow, Poland, provided by the Polish Ministry of Education and Science.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We used the following previously published reference genome assemblies available at NCBI: carrot (GenBank acc. no. GCA_001625215.1), celery (GCA_009905375.1), pepper (GCA_000710875.1), coffee (GCA_003713225.1), fennel (GCA_003724115.1), common sunflower (GCA_002127325.1), Java waterdropwort (GCA_008931105.1), tomato (GCA_000188115.3), potato (GCA_000226075.1), and carrot RNAseq reads from twenty DH1 tissues (PRJNA291977).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
- Bureau, T.E.; Wessler, S.R. Tourist: A large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **1992**, *4*, 1283–1294. [[CrossRef](#)]
- Wessler, S.R.; Bureau, T.E.; White, S.E. LTR-retro-transposons and MITEs: Important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **1995**, *5*, 814–821. [[CrossRef](#)]
- Yang, G.; Hall, T.C. MDM-1 and MDM-2: Two mutator-derived MITE families in rice. *J. Mol. Evol.* **2003**, *56*, 255–264. [[CrossRef](#)]
- Moreno-Vazquez, S.; Ning, J.; Meyers, B.C. *hATpin*, a family of MITE-like *hAT* mobile elements conserved in diverse plant species that forms highly stable secondary structures. *Plant Mol. Biol.* **2005**, *58*, 869–886. [[CrossRef](#)] [[PubMed](#)]
- Nishihara, M.; Hikage, T.; Yamada, E.; Nakatsuka, T. A single-base substitution suppresses flower color mutation caused by a novel miniature inverted-repeat transposable element in gentian. *Mol. Genet. Genom.* **2011**, *286*, 371–382. [[CrossRef](#)]
- Momose, M.; Abe, Y.; Ozeki, Y. Miniature inverted-repeat transposable elements of *Stowaway* are active in potato. *Genetics* **2010**, *186*, 59–66. [[CrossRef](#)]
- Wu, H.; Noda, N.; Mikami, R.; Kang, X.; Akita, Y. Insertion of a novel transposable element disrupts the function of an anthocyanin biosynthesis-related gene in *Echinacea purpurea*. *Sci. Hortic-Amst.* **2021**, *282*, 110021. [[CrossRef](#)]
- Lu, C.; Chen, J.; Zhang, Y.; Hu, Q.; Su, W.; Kuang, H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol. Biol. Evol.* **2012**, *29*, 1005–1017. [[CrossRef](#)] [[PubMed](#)]
- Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71. [[CrossRef](#)]
- Song, X.; Cao, X. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Curr. Opin. Plant Biol.* **2017**, *36*, 111–118. [[CrossRef](#)]
- 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. [[CrossRef](#)] [[PubMed](#)]
- Xu, L.; Yuan, K.; Yuan, M.; Meng, X.; Chen, M.; Wu, J.; Li, J.; Qi, Y. Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements. *Mol. Plant* **2020**, *13*, 851–863. [[CrossRef](#)]
- Ariel, F.D.; Manavella, P.A. When junk DNA turns functional: Transposon-derived non-coding RNAs in plants. *J. Exp. Bot.* **2021**, *72*, 4132–4143. [[CrossRef](#)] [[PubMed](#)]
- Kuang, H.; Padmanabhan, C.; Li, F.; Kamei, A.; Bhaskar, P.B.; Ouyang, S.; Jiang, J.; Buell, C.R.; Baker, B. Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: New functional implications for MITEs. *Genome Res.* **2009**, *19*, 42–56. [[CrossRef](#)]
- Yang, Q.; Li, Z.; Li, W.; Ku, L.; Wang, C.; Ye, J.; Li, K.; Yang, N.; Li, Y.; Zhong, T.; et al. CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16969–16974. [[CrossRef](#)] [[PubMed](#)]
- Mao, H.; Wang, H.; Liu, S.; Li, Z.; Yang, X.; Yan, J.; Li, J.; Tran, L.-S.P.; Qin, F. A transposable element in a *NAC* gene is associated with drought tolerance in maize seedlings. *Nat. Commun.* **2015**, *6*, 8326. [[CrossRef](#)] [[PubMed](#)]
- Henaff, E.; Vives, C.; Desvoves, B.; Chaurasia, A.; Payet, J.; Gutierrez, C.; Casacuberta, J.M. Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of *Brassica* species. *Plant J.* **2014**, *77*, 852–862. [[CrossRef](#)]

19. Morata, J.; Marín, F.; Payet, J.; Casacuberta, J.M. Plant lineage-specific amplification of transcription factor binding motifs by Miniature Inverted-repeat Transposable Elements (MITEs). *Genome Biol. Evol.* **2018**, *10*, 1210–1220. [[CrossRef](#)] [[PubMed](#)]
20. Xin, Y.; Ma, B.; Xiang, Z.; He, N. Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (*Morus notabilis*). *Mob. DNA* **2019**, *10*, 27. [[CrossRef](#)]
21. Oki, N.; Yano, K.; Okumoto, Y.; Tsukiyama, T.; Teraishi, M.; Tanisaka, T. A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. japonica. *Genes Genet. Sys.* **2008**, *83*, 321–329. [[CrossRef](#)] [[PubMed](#)]
22. Niu, X.M.; Xu, Y.C.; Li, Z.W.; Bian, Y.T.; Hou, X.H.; Chen, J.F.; Zou, Y.-P.; Jiang, J.; Wu, Q.; Ge, S.; et al. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6908–6913. [[CrossRef](#)] [[PubMed](#)]
23. Shen, J.; Liu, J.; Xie, K.; Xing, F.; Xiong, F.; Xiao, J.; Li, X.; Xiong, L. Translational repression by a miniature inverted-repeat transposable element in the 3' untranslated region. *Nat. Commun.* **2017**, *8*, 14651. [[CrossRef](#)] [[PubMed](#)]
24. Naito, K.; Zhang, F.; Tsukiyama, T.; Saito, H.; Hancock, C.N.; Richardson, A.O.; Okumoto, Y.; Tanisaka, T.; Wessler, S.R. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **2009**, *461*, 1130–1134. [[CrossRef](#)]
25. Li, J.; Liu, C.; Zhao, A.; Yu, M.; Liu, X.; Chen, X.; Li, Y. A MITE Insertion in the Promoter Region of Anthocyanidin Synthase from *Morus alba* L. *Plant Mol. Biol. Rep.* **2018**, *36*, 188–194. [[CrossRef](#)]
26. Chen, J.; Hu, Q.; Zhang, Y.; Lu, C.; Kuang, H. P-MITE: A database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res.* **2014**, *42*, D1176–D1181. [[CrossRef](#)] [[PubMed](#)]
27. Benjak, A.; Forneck, A.; Casacuberta, J.M. Genome-wide analysis of the “cut-and-paste” transposons of grapevine. *PLoS ONE* **2008**, *3*, e3107. [[CrossRef](#)]
28. Benjak, A.; Boué, S.; Forneck, A.; Casacuberta, J.M. Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). *Genome Biol. Evol.* **2009**, *1*, 75–84. [[CrossRef](#)] [[PubMed](#)]
29. Arscott, S.; Tanumihardjo, S. Carrots of many colors provide basic nutrition and bioavailable phytochemicals acting as a functional food. *Compr. Rev. Food Sci. Food Saf.* **2010**, *9*, 223–239. [[CrossRef](#)]
30. Iorizzo, M.; Ellison, S.; Senalik, D.; Zeng, P.; Satapoomin, P.; Huang, J.; Bowman, M.; Iovene, M.; Sanseverino, W.; Cavagnaro, P.; et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **2016**, *48*, 657. [[CrossRef](#)]
31. Grzebelus, D.; Yau, Y.Y.; Simon, P.W. Master: A novel family of PIF/Harbinger-like transposable elements identified in carrot (*Daucus carota* L.). *Mol. Genet. Genom.* **2006**, *275*, 450–459. [[CrossRef](#)]
32. Macko-Podgórní, A.; Nowicka, A.; Grzebelus, E.; Simon, P.W.; Grzebelus, D. DcSto: Carrot *Stowaway*-like elements are abundant, diverse, and polymorphic. *Genetica* **2013**, *141*, 255–267. [[CrossRef](#)] [[PubMed](#)]
33. Macko-Podgórní, A.; Stelmach, K.; Kwolek, K.; Grzebelus, D. *Stowaway* miniature inverted repeat transposable elements are important agents driving recent genomic diversity in wild and cultivated carrot. *Mob. DNA* **2019**, *10*, 47. [[CrossRef](#)] [[PubMed](#)]
34. Grzebelus, D.; Jagosz, B.; Simon, P.W. The DcMaster Transposon Display maps polymorphic insertion sites in the carrot (*Daucus carota* L.) genome. *Gene* **2007**, *390*, 67–74. [[CrossRef](#)] [[PubMed](#)]
35. Stelmach, K.; Macko-Podgórní, A.; Machaj, G.; Grzebelus, D. Miniature inverted repeat transposable element insertions provide a source of intron length polymorphism markers in the carrot (*Daucus carota* L.). *Front. Plant Sci.* **2017**, *8*, 725. [[CrossRef](#)]
36. Stelmach, K.; Macko-Podgórní, A.; Allender, C.; Grzebelus, D. Genetic diversity structure of western-type carrots. *BMC Plant Biol.* **2021**, *21*, 200. [[CrossRef](#)]
37. Han, Y.; Wessler, S.R. MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **2010**, *38*, e199. [[CrossRef](#)]
38. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
39. Stawujak, K.; Startek, M.; Gambin, A.; Grzebelus, D. MuTANt: A family of Mutator-like transposable elements targeting TA microsatellites in *Medicago truncatula*. *Genetica* **2015**, *143*, 433–440. [[CrossRef](#)]
40. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.R-project.org/> (accessed on 26 May 2020).
41. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
42. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
43. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)]
44. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)]
45. Machaj, G.; Bostan, H.; Macko-Podgórní, A.; Iorizzo, M.; Grzebelus, D. Comparative transcriptomics of root development in wild and cultivated carrots. *Genes* **2018**, *9*, 431. [[CrossRef](#)] [[PubMed](#)]
46. Hartley, S.W.; Mullikin, J.C. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* **2016**, *44*, e127. [[CrossRef](#)]

47. Song, X.; Sun, P.; Yuan, J.; Gong, K.; Li, N.; Meng, F.; Zhang, Z.; Li, X.; Hu, J.; Wang, J.; et al. The celery genome sequence reveals sequential paleo-polyploidizations, karyotype evolution and resistance gene reduction in apiales. *Plant Biotechnol. J.* **2021**, *19*, 731–744. [[CrossRef](#)] [[PubMed](#)]
48. Hulse-Kemp, A.M.; Maheshwari, S.; Stoffel, K.; Hill, T.A.; Jaffe, D.; Williams, S.R.; Van Deynze, A. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* **2018**, *5*, 4. [[CrossRef](#)]
49. Tran, H.T.; Ramaraj, T.; Furtado, A.; Lee, L.S.; Henry, R.J. Use of a draft genome of coffee (*Coffea arabica*) to identify SNP s associated with caffeine content. *Plant Biotechnol. J.* **2018**, *16*, 1756–1766. [[CrossRef](#)] [[PubMed](#)]
50. Palumbo, F.; Galla, G.; Vitulo, N.; Barcaccia, G. First draft genome sequencing of fennel (*Foeniculum Vulgare* Mill.): Identification of simple sequence repeats and their application in marker-assisted breeding. *Mol. Breed.* **2018**, *38*, 122. [[CrossRef](#)]
51. Staton, S.E.; Bakken, B.H.; Blackman, B.K.; Chapman, M.A.; Kane, N.C.; Tang, S.; Ungerer, M.C.; Knapp, S.J.; Rieseberg, L.H.; Burke, J.M. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* **2012**, *72*, 142–153. [[CrossRef](#)]
52. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **2012**, *485*, 635. [[CrossRef](#)]
53. Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*, 189. [[CrossRef](#)] [[PubMed](#)]
54. Darzentas, N. Circoletto: Visualizing sequence similarity with Circos. *Bioinformatics* **2010**, *26*, 2620–2621. [[CrossRef](#)]
55. Akagi, H.; Yokozeki, Y.; Inagaki, A.; Mori, K.; Fujimura, T. Micron, a microsatellite-targeting transposable element in the rice genome. *Mol. Genet. Genom.* **2001**, *266*, 471–480. [[CrossRef](#)] [[PubMed](#)]
56. Franco, J.A.V.; Wang, Y.; Huo, N.; Ponciano, G.; Colvin, H.A.; McMahan, C.M.; Gu, Y.Q.; Belknap, W.R. Modular assembly of transposable element arrays by microsatellite targeting in the guayule and rice genomes. *BMC Genom.* **2018**, *19*, 271. [[CrossRef](#)]
57. Domínguez, M.; Dugas, E.; Benchouaia, M.; Leduque, B.; Jiménez-Gómez, J.M.; Colot, V.; Quadrana, L. The impact of transposable elements on tomato diversity. *Nat. Commun.* **2020**, *11*, 4058. [[CrossRef](#)]
58. Wen, J.; Yu, Y.; Xie, D.F.; Peng, C.; Liu, Q.; Zhou, S.D.; He, X.J. A transcriptome-based study on the phylogeny and evolution of the taxonomically controversial subfamily Apioideae (Apiaceae). *Ann. Bot.-Lond.* **2020**, *125*, 937–953. [[CrossRef](#)] [[PubMed](#)]
59. Batista, R.A.; Moreno-Romero, J.; Qiu, Y.; Van Boven, J.; Santos-González, J.; Figueiredo, D.D.; Köhler, C. The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife* **2019**, *8*, e50541. [[CrossRef](#)]
60. Baud, A.; Wan, M.; Nouaud, D.; Francillonne, N.; Anxolabéhère, D.; Quesneville, H. Traces of transposable elements in genome dark matter co-opted by flowering gene regulation networks. *bioRxiv* **2020**. [[CrossRef](#)]
61. Fortune, P.M.; Roulin, A.; Panaud, O. Horizontal transfer of transposable elements in plants. *Commun. Integr. Biol.* **2008**, *1*, 74–77. [[CrossRef](#)]
62. Zhang, H.H.; Zhou, Q.Z.; Wang, P.L.; Xiong, X.M.; Luchetti, A.; Raoult, D.; Levasseur, A.; Santini, S.; Abergel, C.; Legendre, M.; et al. Unexpected invasion of miniature inverted-repeat transposable elements in viral genomes. *Mob. DNA* **2018**, *9*, 19. [[CrossRef](#)] [[PubMed](#)]
63. Hou, F.; Ma, B.; Xin, Y.; Kuang, L.; He, N. Horizontal transfers of LTR retrotransposons in seven species of Rosales. *Genome* **2018**, *61*, 587–594. [[CrossRef](#)] [[PubMed](#)]
64. De Melo, E.S.; Wallau, G.L. Mosquito genomes are frequently invaded by transposable elements through horizontal transfer. *PLoS Genet.* **2020**, *16*, e1008946. [[CrossRef](#)]
65. Castanera, R.; Vendrell-Mir, P.; Bardil, A.; Carpentier, M.C.; Panaud, O.; Casacuberta, J.M. The amplification dynamics of MITEs and their impact on rice trait variability. *Plant J.* **2021**. [[CrossRef](#)] [[PubMed](#)]