

Article

Cannabis sativa L. Miniature Inverted-Repeat Transposable-Element Landscapes in Wild-Type (JL) and Domesticated Genome (CBDRx)

Mariana Quiroga ^{1,†}, Clara Crociara ^{1,†}, Esteban Schenfeld ^{1,2}, Franco Daniel Fernández ^{2,3}, Juan Crescente ^{2,4}, Leonardo Vanzetti ^{2,4,*} and Marcelo Helguera ^{1,*}

- ¹ Unidad de Estudios Agropecuarios (UDEA) Instituto Nacional de Tecnología Agropecuaria (INTA)—Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Fisiología y Recursos Genéticos Vegetales (IFRGV), Centro de Investigaciones Agropecuarias (CIAP)—INTA, Córdoba X5014MGO, Argentina; quiroga.mariana@inta.gob.ar (M.Q.); crociara.clara@inta.gob.ar (C.C.); schenfeld.esteban@inta.gob.ar (E.S.)
- ² CONICET, Buenos Aires C1425 CABA, Argentina; fernandez.franco@inta.gob.ar (F.D.F.); juan.crescente@gmail.com (J.C.)
- ³ Unidad de Fitopatología y Modelización Agrícola (UFYMA) INTA—CONICET, Instituto de Patología Vegetal (IPAVE), CIAP—INTA, Córdoba X5014MGO, Argentina
- ⁴ Estación Experimental Agropecuaria (EEA) Marcos Juárez—INTA, Marcos Juárez 2580, Argentina
- * Correspondence: vanzetti.leonardo@inta.gob.ar (L.V.); helguera.marcelo@inta.gob.ar (M.H.); Tel.: +54-9-11-6562-9490 (L.V.); +54-9-11-3906-9072 (M.H.)
- † These authors contributed equally to this work.

Abstract: *Cannabis sativa* L. is a globally cultivated plant with significant industrial, nutritional, and medicinal value. Its genome, comprising nine autosomes and sex chromosomes (X and Y), has been extensively studied, particularly in the context of precise breeding for specific enduses. Recent advances have facilitated genome-wide analyses through platforms like the NCBI Comparative Genome Viewer (CGV) and CannabisGDB, among others, enabling comparative studies across multiple Cannabis genotypes. Despite the abundance of genomic data, a particular group of transposable elements, known as miniature inverted-repeat transposable elements (MITEs), remains underexplored in Cannabis. These elements are non-autonomous class II DNA transposons characterized by high copy numbers and insertion preference in non-coding regions, potentially affecting gene expression. In the present study, we report the sequence annotation of MITEs in wild-type and domesticated Cannabis genomes obtained using the MITE Tracker software. We also develop a simple and innovative protocol to identify genome-specific MITE families, offering valuable tools for future research on marker development focused on important genetic variation for breeding in *Cannabis sativa*.

Keywords: Cannabis; MITEs; wild and domesticated genomes



Academic Editor: Mohammad Anwar Hossain

Received: 24 January 2025

Revised: 10 March 2025

Accepted: 19 March 2025

Published: 25 March 2025

Citation: Quiroga, M.; Crociara, C.; Schenfeld, E.; Fernández, F.D.; Crescente, J.; Vanzetti, L.; Helguera, M. *Cannabis sativa* L. Miniature Inverted-Repeat Transposable-Element Landscapes in Wild-Type (JL) and Domesticated Genome (CBDRx). *Int. J. Plant Biol.* **2025**, *16*, 40. <https://doi.org/10.3390/ijpb16020040>

Copyright: © 2025 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

Cannabis (*Cannabis sativa* L.) is an erect annual herb belonging to the family *Cannabaceae*, with a dioecious breeding system [1]. It has a diploid genome ($2n = 20$) with a karyotype composed of nine autosomes and a pair of sex chromosomes (X and Y) [1,2]. From its centre of domestication in Central Asia, Cannabis has spread worldwide. It has been cultivated for millennia either as hemp, for fiber and grain, or for its cannabinoid synthesis for medicinal or recreational purposes [3,4]. Cannabis has considerable industrial, nutritional, and medical value, and information about its genome is extensive and valuable for precise breeding considering previously described specific enduses.

A sequencing milestone was obtaining the first Cannabis draft whole genome covering 534 Mb generated from Purple Kush (PK), a drug-type strain widely used for its medicinal effects [1]. Recently, online platforms such as the NCBI Comparative Genome Viewer (CGV) [5] and CannabisGDB [6], among others, have significantly facilitated genome-wide analyses considering several Cannabis genotypes. Regarding CVG, it offers a user-friendly environment to compare up to 15 different cannabis genotype genomes, from chromosome to intra chromosome and gene region scales, including drug-type cultivars Purple Kush [1], Jamaican Lion DASH [7], hemp-type cultivars Cannbio2 [8], Finola [1], Pink Pepper, Abacus, and feral- or wild-type genotype JL [9]. CannabisGDB is a comprehensive genomics database covering Cannabis genotypes Purple Kush, Jamaican Lion DASH, Finola, CBDRx, Pineapple Banana Bubba Kush, LA Confidential, Chemdog91, and Cannatonic, with genomic tools to analyze and compare at the variety, gene locus, metabolite, and protein levels [5].

Both platforms offer valuable tools for understanding the evolution of the Cannabis genome and how diversity can be generated within this species. Genome-level genetic variation or diversity analysis is a compelling tool for shedding light on the driving forces behind processes such as domestication, breeding, and adaptation in animals and plants [9], including Cannabis.

An interesting source of genetic variation which has not been studied extensively within the Cannabis genome is a particular group of transposable elements (TEs) [10–14] known as miniature inverted-repeat transposable elements (MITEs) [15–18]. There are two classes of TEs, classified according to whether their transposition intermediate is RNA (class 1) or DNA (class 2). Also, each group of TEs contains autonomous and non-autonomous elements depending on the presence or absence of coding regions associated with transposition [10,19]. In class two, MITEs are DNA transposon nonautonomous elements, small in length—typically 70 bp to 600 bp—with well-defined structural features—terminal inverted repeats (TIRs) longer than 10 bp flanked by small direct repeats (target site duplication, TSD, 2–10 bp long) at both ends of the element [10,20]. These TEs are frequently found in or near plant genes [15,21–24], and evidence of these elements potentially altering gene expression in plants has been described [22,25–28].

Another characteristic of MITEs is their very high copy numbers per genome compared to their related autonomous class II and class I transposons [10,29]. Their success in proliferating within the host genome is mainly due to their insertion preference for non-coding regions (promoters, introns, and untranslated regions), which leads to the generation of stable or neutral mutations while avoiding disruption of the exonic regions of genes [29]. MITE-generated mutations in non-coding regions near genes can be used as a potential molecular marker system in plant genetics and breeding [27,30–33] and to understand evolution processes like domestication [29,34,35].

As the structural features defining MITEs are well-characterized, computational approaches have been employed to identify and annotate MITE sequences within genomes [20,24,36–38]. With different algorithms to discover MITEs, these programs exhibit advantages and disadvantages concerning speed, memory use efficiency, and false-positive scoring, among other parameters [24,30], positioning DetectMITE [36] as one of the most efficient, precise, and comprehensive tools in detecting MITEs and MITE Tracker [24] as the most accurate and meticulous in the filtering of false positives [39].

In the present study, the sequence annotation of MITEs in wild-type and domesticated Cannabis genomes is reported. Based on this information, MITE order and organization at the chromosome level in wild and domesticated Cannabis genomes were analyzed. The development of a simple and innovative protocol to identify genome-specific MITE families

offers valuable tools for future research on marker development focused on valuable genetic variation for breeding in Cannabis.

2. Materials and Methods

2.1. Discovery, Annotation, and Organization of Cannabis MITEs

The Cannabis genome assemblies of the wild-type variety JL, deposited at NCBI under Bioproject number PRJNA562042 [40], and of the high-CBD hemp cultivar CBDRx, GenBank assembly accession GCA_900626175.2 [41], were used for MITE discovery and annotation.

The software MITE Tracker [24] was used to find and annotate full MITEs and cluster them into high-sequence homology families using the following parameters: TSD element length: 2–10 nucleotides (nt), MITE length: 50–650 nt, and copy number threshold = 3, with 95–95 homology and coverage.

Chromosomes of the JL and CBDRx genomes were organized based on synteny using the main view of the CGV (“ideogram view”) [4], where pairwise alignments served as connectors that linked syntenic regions between the two assemblies at the chromosomal level. The identified pairs of chromosomes with syntenic regions were then used to determine the MITE content and physical organization, considering chromosome arms and telomeric and centromeric regions.

2.2. Genome-Specific MITE Identification

For genome-specific MITE identification, the software MITE Tracker [24] was run using the previous parameters, including, in this case, both the JL and CBDRx genomes in a single run. This approach allowed us to establish MITE families by identifying highly homologous (more than 95%) MITE members (at least three by default) previously defined in chromosomes from both genomes. With this strategy, we identified MITE families with members occurring exclusively in the JL or CBDRx genomes or both genomes.

2.3. Potential Uses of Genome-Specific MITE for Fingerprinting and Molecular Marker Development

Individual NCBI BLAST searches at <https://blast.ncbi.nlm.nih.gov/Blast.cgi> were conducted using subsets of genome-specific MITEs from JL and CBDRx and random MITEs selected from each JL and CBDRx genome (single run). The searches were against the “whole-genome shotgun contigs” database, restricted to the organism “*Cannabis sativa* (taxid:3483)” and optimized for highly similar sequences (megablast). Four distinct datasets were obtained using the second-highest identity value of each MITE BLAST, as the first identity value was consistently the original JL or CBDRx sequence. After that, an ANOVA test was performed to explore significant differences among groups using the F statistic from the *scipy.stats* library [42,43] with a *p*-value < 0.05. Regarding significant differences between datasets, pairwise differences between groups were identified using Tukey’s HSD with FWER = 0.05.

To explore the potential use of CBDRx genome-specific MITEs in developing molecular markers for physically linked traits, the nearest genes of these MITEs were searched through BLAST searches against the “refseq_reference_genomes” database, restricted to the organism “*Cannabis sativa* (taxid:3483)”. After identifying the nearest genes for each MITE using Genome Data Viewer, the functions of the proteins encoded by these genes were examined through BlastKOALA [44]. To achieve this, the nearest genes to CBDRx genome-specific MITEs were filtered by removing duplicates and gene codes lacking the corresponding protein-coding sequences in the reference Cannabis genome Pink Pepper ASM2916894v1 (GCF_029168945.1). The protein sequences of the remaining genes associated with MITEs were then analyzed using the BlastKOALA platform.

3. Results

3.1. Discovery, Annotation, and Organization of Cannabis MITEs

The densities of MITEs at both the whole-genome and chromosome scales were consistently higher in JL than CBDRx (Table 1). Using MITE Tracker on the JL genome, 691 families were identified, encompassing 14,444 MITEs (File S1). Chromosome CM022971.1 exhibited the highest density of 23 MITEs/Mb, while chromosomes CM022969.1, CM022967.1, CM022968.1, and CM022973.1 showed the lowest density, each with 15 MITEs/Mb (Table 1).

Table 1. Frequency of MITEs (number of MITEs per Mb) in the chromosomes of *Cannabis sativa* JL and CBDRx genomes.

| CBDRx | | | | JL | | | |
|----------------|--------|-------|----------|----------------|--------|-------|-------------------|
| Chr. Accession | Mb | MITEs | MITEs/Mb | Chr. Accession | Mb | MITEs | MITEs/Mb |
| NC_044371.1 | 101.21 | 1394 | 13 | CM022971.1 | 80.62 | 1934 | 23 |
| NC_044375.1 | 96.35 | 998 | 10 | CM022969.1 | 83.00 | 1327 | 15 |
| NC_044372.1 | 94.67 | 997 | 10 | CM022967.1 | 89.82 | 1370 | 15 |
| NC_044373.1 | 91.91 | 1201 | 13 | CM022965.1 | 93.00 | 1612 | 17 |
| NC_044374.1 | 88.18 | 1050 | 11 | CM022968.1 | 83.22 | 1320 | 15 |
| NC_044377.1 | 79.34 | 1099 | 13 | CM022970.1 | 82.47 | 1334 | 16 |
| NC_044378.1 | 71.24 | 844 | 11 | CM022973.1 | 69.09 | 1104 | 15 |
| NC_044379.1 | 64.62 | 929 | 14 | CM022974.1 | 54.53 | 1192 | 21 |
| NC_044376.1 | 61.56 | 861 | 13 | CM022972.1 | 70.97 | 1177 | 16 |
| NC_044370.1 | 104.99 | 1423 | 13 | CM022966.1 | 91.28 | 1791 | 19 |
| Totals | 854.49 | 10903 | 12.1 | | 807.90 | 14444 | 17.2 ¹ |

¹ The data include chromosome sizes (Mb) and the total number of MITEs identified in each chromosome.

In comparison, the CBDRx genome yielded 649 families and 10,903 MITEs (File S2), with chromosome NC_044379.1 exhibiting the highest density of 14 MITEs/Mb (Table 1).

3.2. Genome-Specific MITE Identification

The whole-genome alignment between JL and CBDRx was the first step in comparing chromosome pairs based on syntenic information (Figure S1). Only chromosomes three (CM022967.1/NC_044372.1) and six (CM022970.1/NC_044377.1) consistently shared syntenic information and chromosome numbering. Inconsistencies were observed between syntenic information and chromosome numbering in the remaining chromosomes (one: CM022965.1/NC_044371.1; two: CM022966.1/NC_044375.1; four: CM022968.1/NC_044373.1; five: CM022969.1/NC_044374.1; seven: CM022971.1/NC_044378.1; eight: CM022972.1/NC_044379.1; nine: CM022973/NC_044376.1; and ten: CM022974.1/NC_044370.1).

The dominant pattern in MITE distribution along the JL chromosomes (Figure 1, in blue) was characterized by the highest frequency values observed within the centromeric region, with lower MITE frequencies observed towards both telomeres (chromosomes CM022965.1, CM022967.1, CM022969.1, CM022970.1, and CM022972.1). The opposite pattern (highest MITE frequencies associated with telomeric regions and lowest MITE frequencies within the centromere) was observed in chromosomes CM022966.1, CM022971.1, and CM022973.1, while different patterns from the previous ones were noted in chromosomes CM022968.1 and CM022974.1.

In contrast to the JL chromosomes, the dominant pattern in MITE distribution along the CBDRx chromosomes (Figure 1, in orange) was characterized by the highest MITE frequencies associated with telomeric regions and the lowest MITE frequencies around the centromere (chromosomes NC_044373.1, NC_033370.1, NC_044372.1, NC_044374.1, NC_044375.1, NC_044377.1, NC_044371.1, NC_044376.1, and NC_044378.1). A second pattern, observed in chromosome NC_044379.1, featured the highest MITE frequencies

in one telomere, progressively decreasing through the centromere and towards the second telomere.



Figure 1. Patterns of MITE distribution in *Cannabis sativa* JL (blue) and CBDRx (orange) chromosomes. Chromosomes are referred to by their sequence accession code, with the chromosome number in parentheses. The vertical axis indicates the cumulative number of MITEs located in each chromosome interval. The horizontal axis represents the chromosome length divided into ten equal sections.

When comparing pairs of syntenic chromosomes, CM022966.1/NC_044374.1, CM02273.1/NC_044378, and CM022971.1/NC_044371.1 were the sole combinations sharing a similar telomeric MITE distribution, while the other syntenic chromosome pairs showed inconsistencies in MITE distribution (Figure 1).

3.3. Finding Genome-Specific MITEs

Including the JL and CBDRx genomes in a single run of MITE Tracker allowed for the identification of a total number of 1024 MITE families based on 31155 MITEs (File S3). Among these families, 979 included MITEs belonging to the CBDRx and JL genomes (Table S1). Twenty-one MITE families were detected exclusively in the JL genome. One of them, family F964, comprised 53 MITEs distributed across the ten JL chromosomes. The other 20 JL genome-specific MITE families were constituted by three to nine MITE elements each. Regarding the CBDRx genome, 24 genome-specific MITE families were detected, with 4 families comprising 19, 18, 16, and 10 MITEs and the remaining 21 families containing 3 to 7 MITEs each (Table S1).

3.4. Potential Uses of Genome-Specific MITEs for Fingerprinting and Molecular Marker Development

To explore the use of genome-specific MITEs for fingerprinting, individual NCBI BLAST searches were managed using subsets of genome-specific MITEs from JL and CBDRx (both genomes in a single run of MITE Tracker). The same was conducted with random MITEs selected from the JL and CBDRx genomes (runs of MITE Tracker with separate genomes). Thus, four distinct datasets were obtained based on the second-highest

identity values, as described in the Materials and Methods. The mean identity values obtained in the genome-specific MITEs (only in JL and only in CBDRx) were consistently lower than those obtained in the random MITEs from individual genomes (Figure 2). Significant differences among groups were obtained ($F = 4.166$, $p = 0.0084$), with significant pairwise differences detected in the combinations CBDRx vs. only in CBDRx ($p = 0.005$) and only in CBDRx vs. only in JL ($p = 0.0133$). The remaining pairwise combinations (JL/only in CBDRx; only in JL/CBDRx; JL/CBDRx; JL/only in JL) showed no differences in NCBI BLAST identity values ($p > 0.05$, Figure 2). Considering the potential of genome-specific MITEs for developing molecular markers for interesting traits, its nearest genes were searched by performing BLAST searches against the CBDRx reference genome. With this approach, after excluding duplications and genes not present in the Cannabis reference proteome, the genes associated with CBDRx genome-specific MITEs were analyzed. A subset of genes associated with CBDRx genome-specific MITEs is available in Table S3.

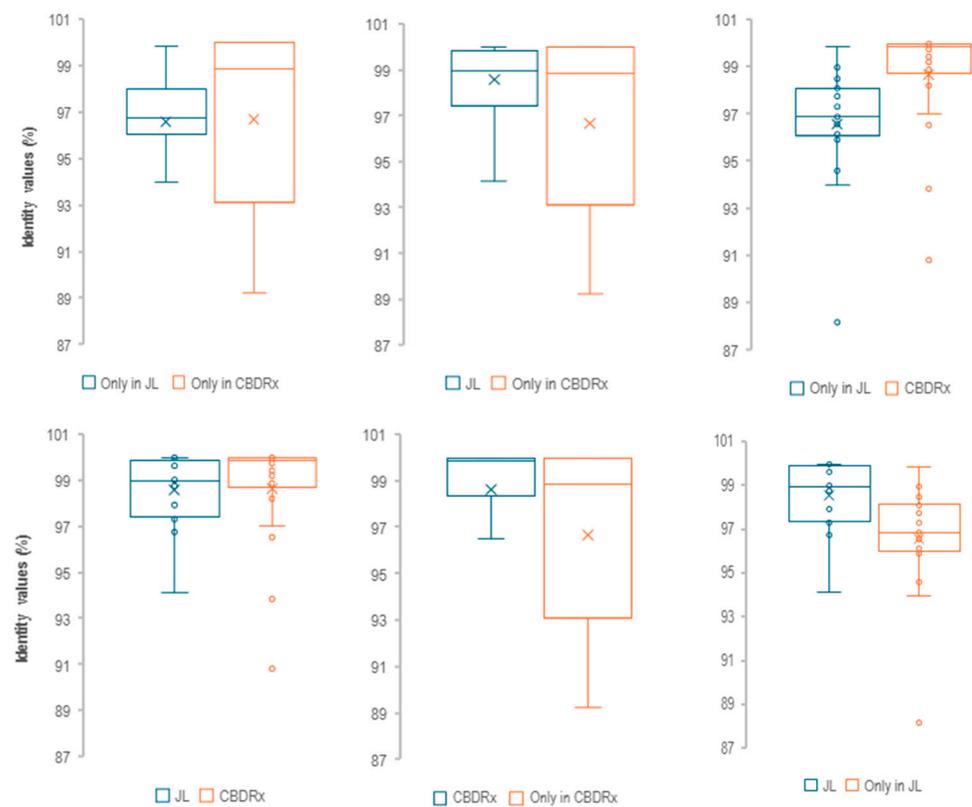


Figure 2. Dispersion of NCBI BLAST second-highest identity values considering four different MITEs datasets. Only in JL: JL MITEs dataset obtained in a single MITE Tracker run including the JL and CBDRx genomes in a single run; only in CBDRx: CBDRx MITEs dataset obtained in a single MITE Tracker run including the JL and CBDRx genomes in a single run; JL, JL MITEs subset obtained in a single MITE Tracker run including exclusively the JL genome; CBDRx, CBDRx MITEs subset obtained in a single MITE Tracker run including exclusively the CBDRx genome. The comparisons CBDRx vs. only in CBDRx and only in CBDRx vs. only in JL showed significant differences ($P < 0.05$, multiple comparison of means: Tukey's HSD, FWER = 0.05); the other comparisons (JL/only in CBDRx; only in JL/CBDRx; JL/CBDRx; JL/only in JL) showed no significant differences ($p > 0.05$).

To investigate the functions of the proteins encoded by these genes, a BlastKOALA search [44] was conducted, resulting in the annotation of 61 entries (48%). Mapping against BRITe hierarchies [45] identified matches with 25 hierarchical groups distributed across three protein families: (i) metabolism (six groups), (ii) genetic information processing (thirteen groups), and (iii) signaling and cellular processes (six groups).

The most represented groups included *ko01000 Enzymes* (twenty-four entries), *ko02000 Transporters* (seven entries), and *ko03400 DNA Repair and Recombination Proteins* (five entries), as shown in Figure 3.

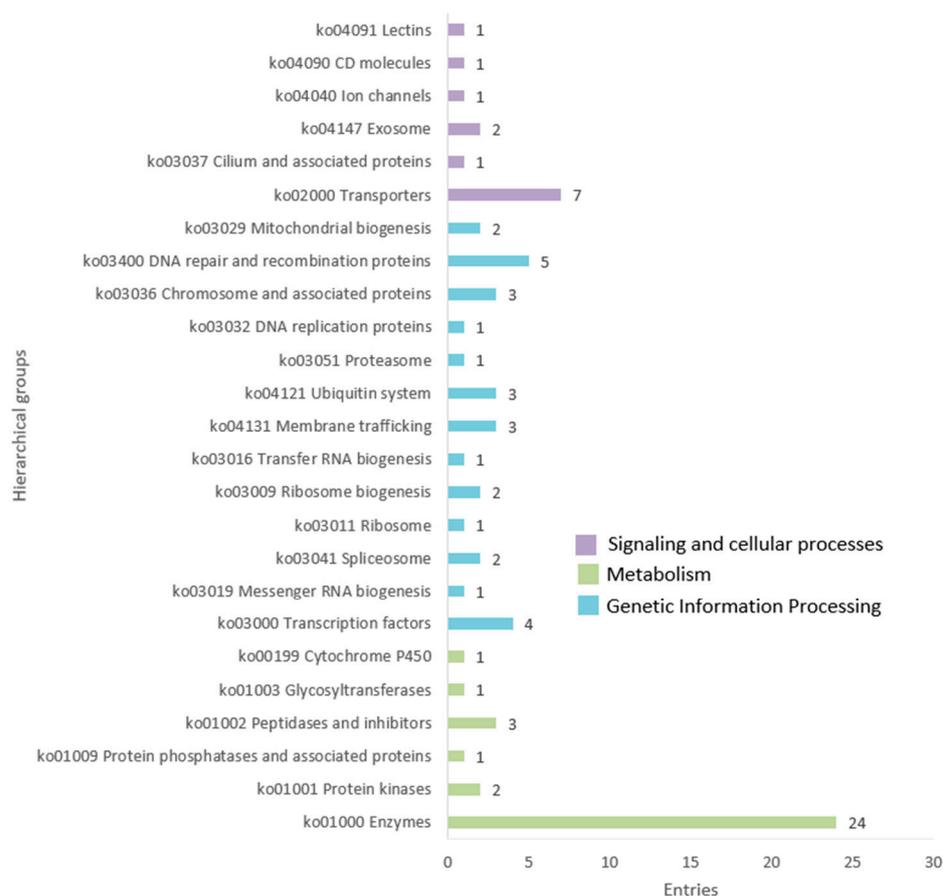


Figure 3. Functional annotation of 127 genes (entries) located near (less than 1000 bp) CBDRx genome-specific MITEs against BRITE hierarchical groups.

4. Discussion

4.1. Density of MITEs Within Cannabis Genomes

In this work, the software MITE Tracker was chosen due to its effectiveness in detecting MITEs concerning false-positive rates and processing efficiency in complex genomes [30]. This is the first time that this program has been used in Cannabis research. Additionally, we proposed a novel and simple method for identifying genome-specific MITEs.

Previous research analyzing MITEs across 38 plant genomes, including Cannabis [46] utilizing the software tools MITE Digger [37], MITE-Hunter [20], and RSPB [22], revealed a wide variation in MITE densities. These ranged from 0.34 MITEs/Mb in *Selaginella moellendorffii* to 480 MITEs/Mb in *Oryza sativa*, with Cannabis exhibiting a density of 140 (Table S2). The MITE density values reported previously [46] were 8 (JL) and 11.5 (CBDRx) times higher than the densities obtained using MITE Tracker in our study (Table 1). Similar patterns were observed in rice, with 140 MITEs/Mb compared to 45.4 MITEs/Mb [24], and in *Vitis vinifera*, with 125 MITEs/Mb compared to 37 MITEs/Mb [47]. This suggests a reduced MITE detection rate in MITE Tracker compared to earlier MITE discovery tools used by Chen et al. [46]. In contrast, Ou et al. [48] reported similar levels of high specificity ($\geq 95\%$) and accuracy ($\geq 94\%$), along with moderate sensitivity (79–81%) but somewhat lower precision (64–79%) when assessing the performance of MITE Tracker and MITE Hunter alongside other MITE discovery tools.

Interestingly, this study revealed a higher density of MITEs in the wild-type and smaller JL genome than the larger, domesticated CBDRx genome (Table 1). This observation contrasts with the positive correlation ($r = 0.72$) between MITE number and genome size reported by Chen et al. [46].

Variation in the number and frequency of MITE copies among closely related species has been reported within the genus *Arabidopsis*. For instance, the genome of *Arabidopsis lyrata* contains 87.88 MITEs/Mb, in contrast to *Arabidopsis thaliana*, which exhibits 27.12 MITEs/Mb [46], see Table S2. A more recent study focused on tea tree (*Camellia sinensis*) modern cultivars also reported variation in MITE copy frequencies considering *C. sinensis* var. *sinensis*, with 11.97 MITEs/Mb, and *C. sinensis* var. *assamica*, with 7.61 [49].

Also, a study on draft genomes of *Triticum* and *Aegilops* species proposed that different levels of MITE proliferation occurred in the A, B, and D subgenomes [50]. For example, the A subgenome exhibited a progressive increase in MITE frequency corresponding to levels of polyploidization but not necessarily related to genome size, as observed in *Triticum urartu* (AA), *T. durum* (AABB), and *T. aestivum* (AABBDD). A similar pattern was observed in the B subgenome for *T. durum* and *T. aestivum*. However, the subgenome D differed, with MITE frequencies being similar in *Aegilops tauschii* (DD) and *T. aestivum* (Table S2). In summary, a contrasting variation in the frequency of MITE copies in bread wheat close relatives *T. urartu* (3.14 MITEs/Mb) and *Ae. tauschii* (6.96 MITEs/Mb) was detected previously [50].

The mechanisms underlying the differing frequencies of MITE proliferation among genetically similar genotypes remain unclear. In our study, a single run of the JL and CBDRx genomes in MITE Tracker allowed us to identify 1024 MITE families, most of which had members in both genomes. Moreover, a significant portion of the differences in MITE proliferation between the JL and CBDRx genomes could be attributed to the ten most successful MITE families in terms of copy number. This small group of families includes 722 (18.7%) JL-specific MITEs (Table S1), contrasting with the 21 families and 134 MITE members detected exclusively in the JL genome when run alone on MITE Tracker. These findings suggest a relaxation in the JL genome of a hypothetical general mechanism regulating MITE multiplication, whereby larger families may have a greater likelihood of proliferating compared to smaller ones.

Regarding MITE distribution within Cannabis chromosomes, two distinct patterns of MITE distribution within syntenic JL and CBDRx chromosomes were observed. Most CBDRx chromosomes exhibited a terminal MITE distribution, characterized by higher MITE densities (MITEs/Mb) in telomeric regions and reduced densities in centromeric (pericentric) areas. In contrast, most JL chromosomes displayed a mostly pericentric distribution, with elevated MITE densities near centromeres and lower values in telomeric regions (Figure 1). Terminal MITE distributions have been previously reported in monocots such as *Brachypodium distachyon*, *Oryza sativa*, and *Sorghum bicolor*, as well as in eudicots like *Vitis vinifera*, *Glycine max*, and *Aquilegia coerulea* [51]. Additional studies describing terminal MITE distributions are also reported in wheat and barley [24,52]. Pericentric MITE distributions have only been documented in the eudicots *Arabidopsis thaliana* and *Brassica rapa* [51]. These authors also observed a strong correlation between MITE terminal distribution in most eudicots and monocots and global gene density due to the general gene/MITE association, interestingly contrasting with the pericentric MITE location, which disagrees with the terminal global gene profile observed in *Brassicaceae*.

A third pattern of MITE distribution is observed in *Citrus* species, where MITE-related sequences are relatively evenly distributed across chromosomes, primarily in gene-flanking regions [53].

Unfortunately, the available JL genome sequence data lacked annotated genes, preventing us from directly comparing our findings with previous data, particularly *Brassicaceae*

genomes. MITEs showed a pericentric distribution in these eudicots' genomes, while genes followed a terminal distribution within chromosomes.

To overcome the inconvenience of lacking annotated genes in the JL genome, we performed a syntenic comparison of CBDRx/JL genome sequences using the NCBI Comparative Genome Viewer (CGV) tool [5] (see Figure S1). The comparisons of the genome sequences revealed, in the first order, inconsistencies in terms of the numbering of chromosomes: only chromosomes three (22967/44372) and six (22970/44377) shared number designations and synteny, and the remaining chromosomes did not. Also, most syntenic chromosome pairs revealed by CGV exhibited major rearrangements as inversions and translocations, including complete chromosomes, e.g., 22,974/44,379, and partial or complete chromosome arms, e.g., 22,965/44,373, 22,966/44,370, 22,967/44,372, 22,968/44,374, 22,969/44,375, 22,970/44,377, and 22,972/44,376. A different situation with conserved syntenic telocentric regions was observed in chromosome pairs 22971/44371 and 22973/44378 (Figure S1).

In agreement with the analysis of this work, comparisons of the JL, cs10 (aka CBDRx), Purple Kush, Finola, and Cannbio-2 genome sequences revealed inconsistencies between these genome sequences regarding orientation and numbering of chromosomes [8]. The comparative analysis carried out by Braich et al. [8] showed a strong synteny between Cannbio-2 and CBDRx, contrasting with the generalized inconsistencies observed in the Cannbio-2 and JL comparison, like what we observed in our study in the CBDRx and JL comparison.

Interestingly, in this study, all cases of JL chromosomes displaying pericentric MITE distributions also showed chromosome arm inversions and/or translocations, such that MITE and gene-rich telomeric regions in CBDRx chromosomes showed synteny with the pericentric region of the corresponding JL chromosome, suggesting a pericentric distribution of MITEs and genes in JL chromosomes, as observed in Figure 4 (see also Figures 1 and S1). The pericentric distributions of MITEs and genes in JL chromosomes with chromosome arm inversions would be an alternative pattern to the terminal (telomeric) distributions of MITEs and genes observed in most eudicots and monocots and pericentric MITE locations combined with the terminal gene distribution described previously for *Brassicaceae* [51]. Moreover, we should be cautious in the hypothesis of pericentric distribution of MITEs and genes within JL chromosomes, as assembly data comparisons [8] would suggest assembly issues. Further sequencing and reassembly of the JL genome will elucidate this issue. The quality of sequence assemblies can vary significantly depending on the characteristics of the initial data. In this case, the JL genome assembly was generated using PacBio and Illumina 350 bp paired-end sequences, which differ from the Nanopore and Illumina 150 bp paired-end sequences used in the CBDRx assembly [40,41]. Subsequently, different heuristic approaches were employed for assembly construction. For the JL genome, tools such as Canu [54], SMARTdenovo [55], QuickMerge [56], and PILON [57] were utilized, followed by BUSCO [58] for genome evaluation and Hi-C [59] for chromosome assembly. The CBDRx assembly involved the use of MINIMAP2 [60], RACON [61], PILON [57], BWA [62], and Hi-C [59]. Additionally, CBDRx incorporated the construction of a genetic linkage map, where contigs were aligned using BWA [62] in combination with SALSAs [63], ALLMAPS [64], RACON [61], and PILON [57]. This step of anchoring the genome sequence to a genetic map represents a key distinction from the JL genome assembly, which did not include this additional anchoring process.

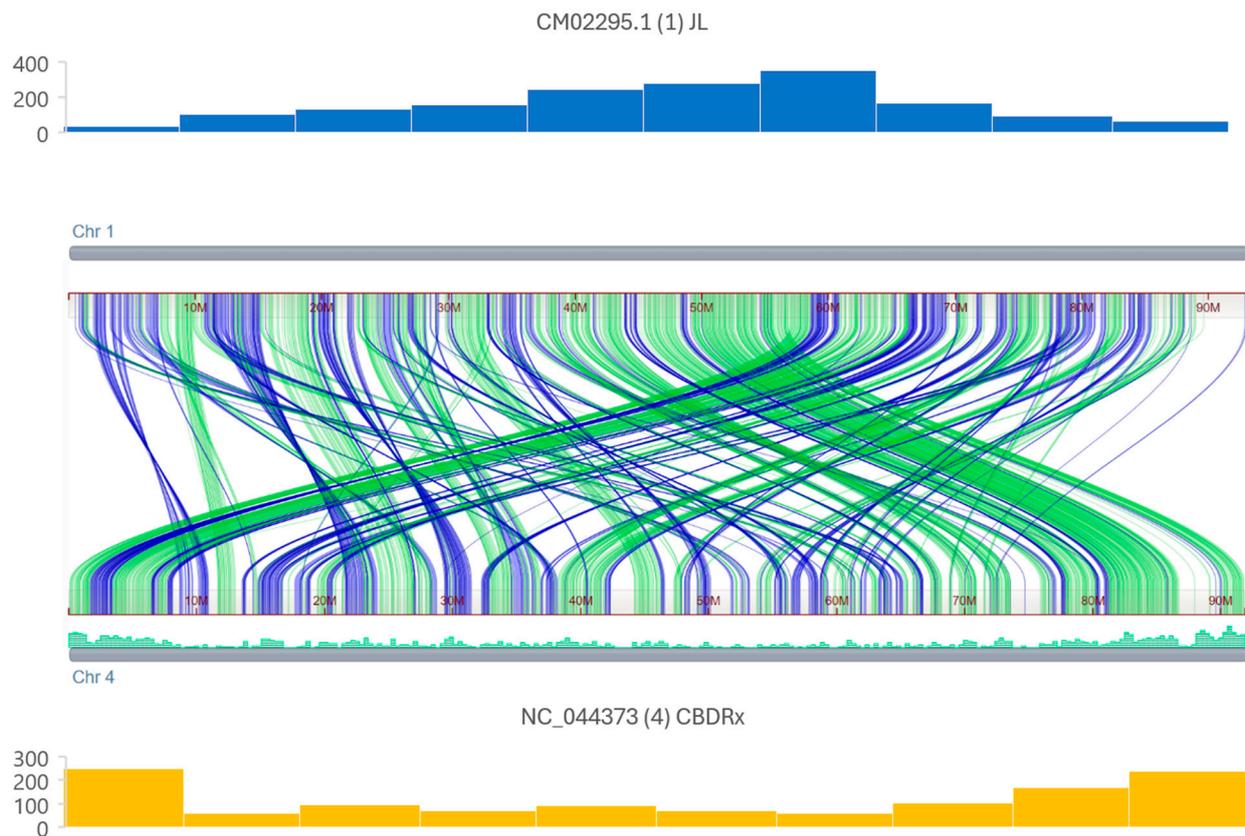


Figure 4. Organization of syntenic regions and MITEs within CM022965.1 and NC_044373 *Cannabis sativa* chromosomes. **In the middle:** syntenic alignment between *Cannabis sativa* JL chromosome CM022965.1 (1) and CBDRx chromosome NC_044373.1 (4). Reciprocal best-placed alignments are shown in light green (forward) and purple (reverse) connectors; minimum alignment size was set in 10,000 bp. In NC_044373, the gene content is also indicated in dark green. **In the top and bottom:** patterns of MITE distribution in the JL chromosome CM022965.1 (blue) and CBDRx chromosome NC_044373.1 (orange). The vertical axis indicates the number of MITEs located in each chromosome interval. The horizontal axis represents the chromosome length divided into ten equal sections.

Although the lack of a gold-standard JL genome sequence somehow limits the robustness of the findings on the pair comparison of the MITE distribution along syntenic chromosomes, the results presented on the MITE annotation and frequency along syntenic chromosomes provide the first insights into these specific transposons within the *Cannabis* genome.

4.2. Potential of SNPs of Genome-Specific MITEs in Fingerprinting and Breeding

In this study, genome-specific MITEs were defined by genome-specific SNPs within the constrained set of the JL and CBDRx genomes included in a single run of MITE Tracker. It was possible to hypothesize that genome-specific MITEs would exhibit a significantly higher number of mutations (and SNPs) than those identified from MITE tracker individual runs. It was expected that there would be lower identity values in genome-specific MITEs compared to ordinary MITEs. In agreement with this argument, consistently lower identity values for genome-specific MITEs, compared to those of randomly selected MITEs from a single genome and estimated to be 2% lower, were observed in this study (CBDRx 98.7%, only in CBDRx 96.7%, JL 98.6%, only in JL 96.6%, see Figure 2).

Some inherent characteristics of MITEs, such as their genomic distribution (often located near or within genes) and sequence structure (the combination of flanking sequences and conserved terminal inverted repeats, TIRs) [24], allow for the design of MITE-specific

primers. These primers facilitate their amplification across different genotypes, enabling dominant marker scoring based on the presence/absence variation in MITEs. Consequently, these DNA structures serve as valuable tools for marker development in various applications, including QTL mapping, molecular breeding, and genetic fingerprinting [30,49,65]. Chang, O'Donoghue, and Bureau [66] demonstrated this using inter-MITE polymorphism for fingerprinting barley cultivars and performing genetic similarity analysis.

A similar approach was employed by Dai et al. [33], where 80 polymorphic markers derived from 55 MITEs were developed and used to evaluate genetic diversity in a panel of *B. napus* accessions, comprising 101 natural and 25 synthetic genotypes. Additionally, two of these markers were anchored to candidate genes associated with agronomic traits. The insertion of TEs into genes that affect various crop traits has been well documented [33]. For instance, Studer et al. [67] showed that inserting a TE in the regulatory region of a maize domestication gene acted as an enhancer of gene expression, partially explaining the increased apical dominance in maize. Moreover, MITEs have a higher potential for gene regulation and phenotypic influence than other types of TEs.

In this study, 41.6% of CBDRx genome-specific MITEs were located within 5000 bp of the nearest gene, 8.6% were found within 500 bp, and three were identified as being inserted into genes. Similarly, Feng [68] identified 18 members of a specific MITE family in annotated rice sequences. Of these, 40% were in introns, 30% were less than 530 bp from a coding sequence, and 20% were situated between 1 and 5 kb from a coding sequence; however, none were found within coding regions.

In this context, the proposed approach to identifying genome-specific MITEs contributes to discovering JL- and CBDRx-genome-specific mutations for marker development. The strategy of using MITE Tracker with two genomes to identify genome-specific MITEs can be extended to various Cannabis genotypes and genome pairs from other crops. Depending on the genome size, breeding or research objectives, and data availability, MITE Tracker could also be used to detect specific MITEs at the chromosome level, considering syntenic chromosomes from more than two genotypes. Further studies are required to validate these findings. Based on available information, MITE Tracker has been utilized exclusively for MITE annotation at the genome level in a diverse range of single genomes, including crops such as wheat and rice [24], *Eragrostis* [69], finger millet [70], Ethiopian mustard [71], white fonio [72], and potato [73]; insects such as moths [74,75] and subterranean termites [76]; and fungi such as the barley-covered smut fungus (*Ustilago hordei*) [77] and *Zymoseptoria tritici* [78], among other organisms. Consequently, the approach of employing MITE Tracker to analyze two genomes simultaneously is both a novel and promising strategy for identifying genome-specific mutations to facilitate marker development for selection and fingerprinting, among other uses.

5. Conclusions

When comparing the JL and CBDRx genomes, quantitative differences in MITE organization (count per Mb) were observed between genomes, with the wild-type JL genome exhibiting higher frequencies than the domesticated CBDRx genome.

Since the genomes of the wild-type variety JL and the hemp cultivar CBDRx were assembled using different heuristics, conclusions about the physical distribution of MITEs remain inconclusive.

Also, the approach of leveraging MITE Tracker to analyze two genomes simultaneously represents a novel and effective strategy for identifying genome-specific MITEs. This methodology facilitated the discovery of JL- and CBDRx-specific mutations, offering valuable insights for marker development, selection, and fingerprinting applications.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijpb16020040/s1>, Figure S1: Whole-genome alignment between *Cannabis sativa* JL and CBDRx; Table S1: JL and CBDRx MITEs organized by families and chromosomes; Table S2. MITE frequencies at genome level in different organisms; Table S3: List of genes near CBDRx genome specific MITEs. File S1: MITEs annotation of JL genome; File S2 MITE annotation of CBDRx genome; File S3: MITE annotation of the JL and CBDRx genomes in a single run of MITE Tracker.

Author Contributions: Conceptualization, M.Q., J.C., L.V. and M.H.; Data curation, C.C., E.S., F.D.F. and M.H.; Formal analysis, C.C., J.C., L.V. and M.H.; Funding acquisition, M.Q. and M.H.; Investigation, M.H.; Methodology, C.C. and J.C.; Project administration, M.H.; Resources, M.H.; Software, M.Q., J.C., L.V. and M.H.; Supervision, L.V.; Validation, M.Q., J.C. and M.H.; Writing—original draft, M.Q., C.C., E.S., J.C., F.D.F., L.V. and M.H.; Writing—review and editing, M.Q., C.C., E.S., J.C., F.D.F., L.V. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by INTA, grant numbers PTi512, PEi071, and PEi084.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: Special thanks to Marina Omelchenko (NIH/NLM/NCBI) for giving us access to the Comparative Genome Viewer (CGV) considering the JL and CBDRx genomes.

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

References

- van Bakel, H.; Stout, J.M.; Cote, A.G.; Tallon, C.M.; Sharpe, A.G.; Hughes, T.R.; Page, J.E. The Draft Genome and Transcriptome of *Cannabis sativa*. *Genome Biol.* **2011**, *12*, R102. [[CrossRef](#)] [[PubMed](#)]
- Sakamoto, K.; Akiyama, Y.; Fukui, K.; Kamada, H.; Satoh, S. Characterization; Genome Sizes and Morphology of Sex Chromosomes in Hemp (*Cannabis sativa* L.). *Cytologia* **1998**, *63*, 459–464. [[CrossRef](#)]
- ElSohly, M.A.; Gul, W. Constituents of *Cannabis sativa*. In *Handbook of Cannabis*; Pertwee, R., Ed.; Oxford University Press: Oxford, UK, 2014; pp. 3–22.
- Clarke, R.C.; Merlin, M.D. Cannabis Domestication, Breeding History, Present-Day Genetic Diversity, and Future Prospects. *CRC Crit. Rev. Plant Sci.* **2016**, *35*, 293–327. [[CrossRef](#)]
- Rangwala, S.H.; Rudnev, D.V.; Ananiev, V.V.; Oh, D.-H.; Asztalos, A.; Benica, B.; Borodin, E.A.; Bouk, N.; Evgeniev, V.I.; Kodali, V.K.; et al. The NCBI Comparative Genome Viewer (CGV) Is an Interactive Visualization Tool for the Analysis of Whole-Genome Eukaryotic Alignments. *PLoS Biol.* **2024**, *22*, e3002405. [[CrossRef](#)]
- Cai, S.; Zhang, Z.; Huang, S.; Bai, X.; Huang, Z.; Zhang, Y.J.; Huang, L.; Tang, W.; Haughn, G.; You, S.; et al. CannabisGDB: A Comprehensive Genomic Database for *Cannabis sativa* L. *Plant Biotechnol. J.* **2021**, *19*, 857–859. [[CrossRef](#)]
- McKernan, K.; Helbert, Y.; Kane, L.T.; Ebling, H.; Zhang, L.; Liu, B.; Eaton, Z.; Sun, L.; Dimalanta, E.T.; Kingan, S.; et al. Cryptocurrencies and Zero Mode Wave Guides: An Unclouded Path to a More Contiguous *Cannabis sativa* L. Genome Assembly. *OSF Prepr.* **2018**, 1–21. [[CrossRef](#)]
- Braich, S.; Baillie, R.C.; Spangenberg, G.C.; Cogan, N.O.I. A New and Improved Genome Sequence of *Cannabis sativa*. *GigaByte* **2020**, *2020*, gigabyte10. [[CrossRef](#)]
- Andersson, L.; Purugganan, M. Molecular Genetic Variation of Animals and Plants under Domestication. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2122150119. [[CrossRef](#)]
- Feschotte, C.; Jiang, N.; Wessler, S.R. Plant Transposable Elements: Where Genetics Meets Genomics. *Nat. Rev. Genet.* **2002**, *3*, 329–341. [[CrossRef](#)]
- Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten Things You Should Know about Transposable Elements. *Genome Biol.* **2018**, *19*, 199. [[CrossRef](#)]
- Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory Activities of Transposable Elements: From Conflicts to Benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)] [[PubMed](#)]
- Pulido, M.; Casacuberta, J.M. Transposable Element Evolution in Plant Genome Ecosystems. *Curr. Opin. Plant Biol.* **2023**, *75*, 102418. [[CrossRef](#)]
- Pandita, D.; Pandita, A. *Plant Transposable Elements*; Apple Academic Press: New York, NY, USA, 2023; ISBN 9781003315193.
- 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](#)] [[PubMed](#)]

16. Fattash, I.; Rooke, R.; Wong, A.; Hui, C.; Luu, T.; Bhardwaj, P.; Yang, G. Miniature Inverted-Repeat Transposable Elements: Discovery, Distribution, and Activity. *Genome* **2013**, *56*, 475–486. [[CrossRef](#)]
17. Jiang, N.; Feschotte, C.; Zhang, X.; Wessler, S.R. Using Rice to Understand the Origin and Amplification of Miniature Inverted Repeat Transposable Elements (MITEs). *Curr. Opin. Plant Biol.* **2004**, *7*, 115–119. [[CrossRef](#)] [[PubMed](#)]
18. Pegler, J.L.; Oultram, J.M.J.; Mann, C.W.G.; Carroll, B.J.; Grof, C.P.L.; Eamens, A.L. Miniature Inverted-Repeat Transposable Elements: Small DNA Transposons That Have Contributed to Plant MICRORNA Gene Evolution. *Plants* **2023**, *12*, 1101. [[CrossRef](#)]
19. 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](#)]
20. 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](#)]
21. Bureau, T.E.; Wessler, S.R. Mobile Inverted-Repeat Elements of the Tourist Family Associated with the Genes of Many Cereal Grasses. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1411–1415. [[CrossRef](#)]
22. 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](#)]
23. 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](#)]
24. Crescente, J.M.; Zavallo, D.; Helguera, M.; Vanzetti, L.S. MITE Tracker: An Accurate Approach to Identify Miniature Inverted-Repeat Transposable Elements in Large Genomes. *BMC Bioinform.* **2018**, *19*, 348. [[CrossRef](#)]
25. 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](#)] [[PubMed](#)]
26. 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](#)]
27. Yin, S.; Wan, M.; Guo, C.; Wang, B.; Li, H.; Li, G.; Tian, Y.; Ge, X.; King, G.J.; Liu, K.; et al. Transposon Insertions within Alleles of BnaFLC.A10 and BnaFLC.A2 Are Associated with Seasonal Crop Type in Rapeseed. *J. Exp. Bot.* **2020**, *71*, 4729–4741. [[CrossRef](#)]
28. Jeong, H.; Yun, Y.B.; Jeong, S.Y.; Cho, Y.; Kim, S. Characterization of Miniature Inverted Repeat Transposable Elements Inserted in the CitRWP Gene Controlling Nucellar Embryony and Development of Molecular Markers for Reliable Genotyping of CitRWP in Citrus Species. *Sci. Hortic.* **2023**, *315*, 112003. [[CrossRef](#)]
29. Naito, K.; Cho, E.; Yang, G.; Campbell, M.A.; Yano, K.; Okumoto, Y.; Tanisaka, T.; Wessler, S.R. Dramatic Amplification of a Rice Transposable Element during Recent Domestication. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17620–17625. [[CrossRef](#)]
30. Venkatesh; Nandini, B. Miniature Inverted-Repeat Transposable Elements (MITEs), Derived Insertional Polymorphism as a Tool of Marker Systems for Molecular Plant Breeding. *Mol. Biol. Rep.* **2020**, *47*, 3155–3167. [[CrossRef](#)]
31. von Zitzewitz, J.; Szűcs, P.; Dubcovsky, J.; Yan, L.; Francia, E.; Pecchioni, N.; Casas, A.; Chen, T.H.H.; Hayes, P.M.; Skinner, J.S. Molecular and Structural Characterization of Barley Vernalization Genes. *Plant Mol. Biol.* **2005**, *59*, 449–467. [[CrossRef](#)]
32. Vaschetto, L.M. Miniature Inverted-Repeat Transposable Elements (MITEs) and Their Effects on the Regulation of Major Genes in Cereal Grass Genomes. *Mol. Breed.* **2016**, *36*, 30. [[CrossRef](#)]
33. Dai, S.; Hou, J.; Qin, M.; Dai, Z.; Jin, X.; Zhao, S.; Dong, Y.; Wang, Y.; Wu, Z.; Lei, Z. Diversity and Association Analysis of Important Agricultural Trait Based on Miniature Inverted-Repeat Transposable Element Specific Marker in *Brassica napus* L. *Oil Crop Sci.* **2021**, *6*, 28–34. [[CrossRef](#)]
34. Poretti, M.; Praz, C.R.; Meile, L.; Kälin, C.; Schaefer, L.K.; Schläfli, M.; Widrig, V.; Sanchez-Vallet, A.; Wicker, T.; Bourras, S. Domestication of High-Copy Transposons Underlays the Wheat Small RNA Response to an Obligate Pathogen. *Mol. Biol. Evol.* **2020**, *37*, 839–848. [[CrossRef](#)]
35. Castanera, R.; Vendrell-Mir, P.; Bardil, A.; Carpentier, M.; Panaud, O.; Casacuberta, J.M. Amplification Dynamics of Miniature Inverted-repeat Transposable Elements and Their Impact on Rice Trait Variability. *Plant J.* **2021**, *107*, 118–135. [[CrossRef](#)]
36. Ye, C.; Ji, G.; Liang, C. DetectMITE: A Novel Approach to Detect Miniature Inverted Repeat Transposable Elements in Genomes. *Sci. Rep.* **2016**, *6*, 19688. [[CrossRef](#)]
37. Yang, G. MITE Digger, an Efficient and Accurate Algorithm for Genome Wide Discovery of Miniature Inverted Repeat Transposable Elements. *BMC Bioinform.* **2013**, *14*, 186. [[CrossRef](#)]
38. Hu, J.; Zheng, Y.; Shang, X. MiteFinder: A Fast Approach to Identify Miniature Inverted-Repeat Transposable Elements on a Genome-Wide Scale. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 164–168.
39. Satovic, E.; Cvitanic, E.T.; Plohl, M. Tools and Databases for Solving Problems in Detection and Identification of Repetitive DNA Sequences. *Period Biol.* **2020**, *121–122*, 7–14. [[CrossRef](#)]

40. Gao, S.; Wang, B.; Xie, S.; Xu, X.; Zhang, J.; Pei, L.; Yu, Y.; Yang, W.; Zhang, Y. A High-Quality Reference Genome of Wild *Cannabis sativa*. *Hortic. Res.* **2020**, *7*, 73. [[CrossRef](#)] [[PubMed](#)]
41. Grassa, C.J.; Weiblen, G.D.; Wenger, J.P.; Dabney, C.; Poplawski, S.G.; Timothy Motley, S.; Michael, T.P.; Schwartz, C.J. A New Cannabis Genome Assembly Associates Elevated Cannabidiol (CBD) with Hemp Introgressed into Marijuana. *New Phytol.* **2021**, *230*, 1665–1679. [[CrossRef](#)]
42. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010.
43. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
44. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [[CrossRef](#)]
45. Kanehisa, M.; Sato, Y. KEGG Mapper for Inferring Cellular Functions from Protein Sequences. *Protein Sci.* **2020**, *29*, 28–35. [[CrossRef](#)] [[PubMed](#)]
46. 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](#)]
47. Onetto, C.A.; Ward, C.M.; Borneman, A.R. The Genome Assembly of *Vitis vinifera* Cv. Shiraz. *Aust. J. Grape Wine Res.* **2023**, *2023*, 6686706. [[CrossRef](#)]
48. Ou, S.; Su, W.; Liao, Y.; Chougule, K.; Agda, J.R.A.; Hellinga, A.J.; Lugo, C.S.B.; Elliott, T.A.; Ware, D.; Peterson, T.; et al. Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline. *Genome Biol.* **2019**, *20*, 275. [[CrossRef](#)] [[PubMed](#)]
49. Rohilla, M.; Mazumder, A.; Saha, D.; Pal, T.; Begam, S.; Mondal, T.K. Genome-Wide Identification and Development of Miniature Inverted-Repeat Transposable Elements and Intron Length Polymorphic Markers in Tea Plant (*Camellia sinensis*). *Sci. Rep.* **2022**, *12*, 16233. [[CrossRef](#)]
50. Keidar-Friedman, D.; Bariah, I.; Kashkush, K. Genome-Wide Analyses of Miniature Inverted-Repeat Transposable Elements Reveals New Insights into the Evolution of the Triticum-Aegilops Group. *PLoS ONE* **2018**, *13*, e0204972. [[CrossRef](#)]
51. Boutanaev, A.M.; Osbourn, A.E. Multigenome Analysis Implicates Miniature Inverted-Repeat Transposable Elements (MITEs) in Metabolic Diversification in Eudicots. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E6650–E6658. [[CrossRef](#)]
52. Li, R.; Yao, J.; Cai, S.; Fu, Y.; Lai, C.; Zhu, X.; Cui, L.; Li, Y. Genome-Wide Characterization and Evolution Analysis of Miniature Inverted-Repeat Transposable Elements in Barley (*Hordeum vulgare*). *Front. Plant Sci.* **2024**, *15*, E6650–E6658. [[CrossRef](#)]
53. Liu, Y.; Tahir ul Qamar, M.; Feng, J.-W.; Ding, Y.; Wang, S.; Wu, G.; Ke, L.; Xu, Q.; Chen, L.-L. Comparative Analysis of Miniature Inverted-Repeat Transposable Elements (MITEs) and Long Terminal Repeat (LTR) Retrotransposons in Six Citrus Species. *BMC Plant Biol.* **2019**, *19*, 140. [[CrossRef](#)]
54. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and Accurate Long-Read Assembly via Adaptive *k*-Mer Weighting and Repeat Separation. *Genome Res.* **2017**, *27*, 722–736. [[CrossRef](#)]
55. Istace, B.; Friedrich, A.; d’Agata, L.; Faye, S.; Payen, E.; Beluche, O.; Caradec, C.; Davidas, S.; Cruaud, C.; Liti, G.; et al. De Novo Assembly and Population Genomic Survey of Natural Yeast Isolates with the Oxford Nanopore MinION Sequencer. *Gigascience* **2017**, *6*, 1–13. [[CrossRef](#)]
56. Chakraborty, M.; Baldwin-Brown, J.G.; Long, A.D.; Emerson, J.J. Contiguous and Accurate *de Novo* Assembly of Metazoan Genomes with Modest Long Read Coverage. *Nucleic Acids Res.* **2016**, *44*, e147. [[CrossRef](#)] [[PubMed](#)]
57. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
58. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
59. Belaghzal, H.; Dekker, J.; Gibcus, J.H. Hi-C 2.0: An Optimized Hi-C Procedure for High-Resolution Genome-Wide Mapping of Chromosome Conformation. *Methods* **2017**, *123*, 56–65. [[CrossRef](#)] [[PubMed](#)]
60. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
61. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and Accurate *de Novo* Genome Assembly from Long Uncorrected Reads. *Genome Res.* **2017**, *27*, 737–746. [[CrossRef](#)] [[PubMed](#)]
62. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
63. Ghurye, J.; Pop, M.; Koren, S.; Bickhart, D.; Chin, C.-S. Scaffolding of Long Read Assemblies Using Long Range Contact Information. *BMC Genom.* **2017**, *18*, 527. [[CrossRef](#)]
64. Tang, H.; Zhang, X.; Miao, C.; Zhang, J.; Ming, R.; Schnable, J.C.; Schnable, P.S.; Lyons, E.; Lu, J. ALLMAPS: Robust Scaffold Ordering Based on Multiple Maps. *Genome Biol.* **2015**, *16*, 3. [[CrossRef](#)]

65. Hadagali, S.; Stelmach-Wityk, K.; Macko-Podgórn, A.; Cholin, S.; Grzebelus, D. Polymorphic Insertions of DcSto Miniature Inverted-Repeat Transposable Elements Reveal Genetic Diversity Structure within the Cultivated Carrot. *J. Appl. Genet.* **2024**, *Online ahead of print*. [[CrossRef](#)]
66. Chang, R.-Y.; O'Donoghue, L.S.; Bureau, T.E. Inter-MITE Polymorphisms (IMP): A High Throughput Transposon-Based Genome Mapping and Fingerprinting Approach. *Theor. Appl. Genet.* **2001**, *102*, 773–781. [[CrossRef](#)]
67. Studer, A.; Zhao, Q.; Ross-Ibarra, J.; Doebley, J. Identification of a Functional Transposon Insertion in the Maize Domestication Gene Tb1. *Nat. Genet.* **2011**, *43*, 1160–1163. [[CrossRef](#)] [[PubMed](#)]
68. Feng, Y. Plant MITES: Useful Tools for Plant Genetics and Genomics. *Genom. Proteom. Bioinform.* **2003**, *1*, 90–100. [[CrossRef](#)]
69. VanBuren, R.; Man Wai, C.; Wang, X.; Pardo, J.; Yocca, A.E.; Wang, H.; Chaluvadi, S.R.; Han, G.; Bryant, D.; Edger, P.P.; et al. Exceptional Subgenome Stability and Functional Divergence in the Allotetraploid Ethiopian Cereal Teff. *Nat. Commun.* **2020**, *11*, 884. [[CrossRef](#)]
70. Devos, K.M.; Qi, P.; Bahri, B.A.; Gimode, D.M.; Jenike, K.; Manthi, S.J.; Lule, D.; Lux, T.; Martinez-Bello, L.; Pendergast, T.H.; et al. Genome Analyses Reveal Population Structure and a Purple Stigma Color Gene Candidate in Finger Millet. *Nat. Commun.* **2023**, *14*, 3694. [[CrossRef](#)] [[PubMed](#)]
71. Yim, W.C.; Swain, M.L.; Ma, D.; An, H.; Bird, K.A.; Curdie, D.D.; Wang, S.; Ham, H.D.; Luzuriaga-Neira, A.; Kirkwood, J.S.; et al. The Final Piece of the Triangle of U: Evolution of the Tetraploid *Brassica Carinata* Genome. *Plant Cell* **2022**, *34*, 4143–4172. [[CrossRef](#)]
72. Wang, X.; Chen, S.; Ma, X.; Yssel, A.E.J.; Chaluvadi, S.R.; Johnson, M.S.; Gangashetty, P.; Hamidou, F.; Sanogo, M.D.; Zwaenepoel, A.; et al. Genome Sequence and Genetic Diversity Analysis of an Under-Domesticated Orphan Crop, White Fonio (*Digitaria exilis*). *Gigascience* **2021**, *10*, giab013. [[CrossRef](#)]
73. Zavallo, D.; Crescente, J.M.; Gantuz, M.; Leone, M.; Vanzetti, L.S.; Masuelli, R.W.; Asurmendi, S. Genomic Re-Assessment of the Transposable Element Landscape of the Potato Genome. *Plant Cell Rep.* **2020**, *39*, 1161–1174. [[CrossRef](#)]
74. Klai, K.; Zidi, M.; Chénais, B.; Denis, F.; Caruso, A.; Casse, N.; Mezghani Khemakhem, M. Miniature Inverted-Repeat Transposable Elements (MITEs) in the Two Lepidopteran Genomes of *Helicoverpa armigera* and *Helicoverpa zea*. *Insects* **2022**, *13*, 313. [[CrossRef](#)]
75. Klai, K.; Chénais, B.; Zidi, M.; Djebbi, S.; Caruso, A.; Denis, F.; Confais, J.; Badawi, M.; Casse, N.; Mezghani Khemakhem, M. Screening of *Helicoverpa armigera* Mobilome Revealed Transposable Element Insertions in Insecticide Resistance Genes. *Insects* **2020**, *11*, 879. [[CrossRef](#)]
76. Martelossi, J.; Forni, G.; Iannello, M.; Savojardo, C.; Martelli, P.L.; Casadio, R.; Mantovani, B.; Luchetti, A.; Rota-Stabelli, O. Wood Feeding and Social Living: Draft Genome of the Subterranean Termite *Reticulitermes lucifugus* (Blattodea; Termitoidae). *Insect Mol. Biol.* **2023**, *32*, 118–131. [[CrossRef](#)] [[PubMed](#)]
77. Depotter, J.R.L.; Ökmen, B.; Ebert, M.K.; Beckers, J.; Kruse, J.; Thines, M.; Doehlemann, G. High Nucleotide Substitution Rates Associated with Retrotransposon Proliferation Drive Dynamic Secretome Evolution in Smut Pathogens. *Microbiol. Spectr.* **2022**, *10*, e0034922. [[CrossRef](#)] [[PubMed](#)]
78. Fouché, S.; Badet, T.; Oggenfuss, U.; Plissonneau, C.; Francisco, C.S.; Croll, D. Stress-Driven Transposable Element De-Repression Dynamics and Virulence Evolution in a Fungal Pathogen. *Mol. Biol. Evol.* **2020**, *37*, 221–239. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.