

RESEARCH

Open Access



# Recent amplification of microsatellite-associated miniature inverted-repeat transposable elements in the pineapple genome

Lianyu Lin<sup>1,2</sup>, Anupma Sharma<sup>1</sup> and Qingyi Yu<sup>1\*</sup>

## Abstract

**Background:** Miniature inverted-repeat transposable elements (MITEs) are non-autonomous DNA transposable elements that play important roles in genome organization and evolution. Genome-wide identification and characterization of MITEs provide essential information for understanding genome structure and evolution.

**Results:** We performed genome-wide identification and characterization of MITEs in the pineapple genome. The top two MITE families, accounting for 29.39% of the total MITEs and 3.86% of the pineapple genome, have insertion preference in (TA)<sub>n</sub> dinucleotide microsatellite regions. We therefore named these MITEs *A. comosus* microsatellite-associated MITEs (Ac-mMITEs). The two Ac-mMITE families, Ac-mMITE-1 and Ac-mMITE-2, shared sequence similarity in the terminal inverted repeat (TIR) regions, suggesting that these two Ac-mMITE families might be derived from a common or closely related autonomous elements. The Ac-mMITEs are frequently clustered via adjacent insertions. Among the 21,994 full-length Ac-mMITEs, 46.1% of them were present in clusters. By analyzing the Ac-mMITEs without (TA)<sub>n</sub> microsatellite flanking sequences, we found that Ac-mMITEs were likely derived from *Mutator*-like DNA transposon. Ac-MITEs showed highly polymorphic insertion sites between cultivated pineapples and their wild relatives. To better understand the evolutionary history of Ac-mMITEs, we filtered and performed comparative analysis on the two distinct groups of Ac-mMITEs, microsatellite-targeting MITEs (mt-MITEs) that are flanked by dinucleotide microsatellites on both sides and mutator-like MITEs (ml-MITEs) that contain 9/10 bp TSDs. Epigenetic analysis revealed a lower level of host-induced silencing on the mt-MITEs in comparison to the ml-MITEs, which partially explained the significantly higher abundance of mt-MITEs in pineapple genome. The mt-MITEs and ml-MITEs exhibited differential insertion preference to gene-related regions and RNA-seq analysis revealed their differential influences on expression regulation of nearby genes.

\* Correspondence: [qyu@ag.tamu.edu](mailto:qyu@ag.tamu.edu)

<sup>1</sup>Texas A&M Agrilife Research Center at Dallas, Texas A&M University System, Dallas, TX 75252, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions:** Ac-mMITEs are the most abundant MITEs in the pineapple genome and they were likely derived from *Mutator*-like DNA transposon. Preferential insertion in (TA)<sub>n</sub> microsatellite regions of Ac-mMITEs occurred recently and is likely the result of damage-limiting strategy adapted by Ac-mMITEs during co-evolution with their host. Insertion in (TA)<sub>n</sub> microsatellite regions might also have promoted the amplification of mt-MITEs. In addition, mt-MITEs showed no or negligible impact on nearby gene expression, which may help them escape genome control and lead to their amplification.

**Keywords:** Pineapple, Miniature inverted-repeat transposable elements, (TA)<sub>n</sub> dinucleotide microsatellite, *Ananas*

## Background

Miniature inverted-repeat transposable elements (MITEs) are non-autonomous DNA transposable elements (TEs), transposing by a “cut and paste” mechanism. MITEs were first described in plant genomes [1] and later found in a wide range of organisms, including invertebrates [2, 3], vertebrates [4], fungi [5], and viruses [6]. MITEs are characterized by a small size (< 500 bp), a high copy number, a stable secondary structure, and terminal inverted repeats (TIRs) flanked by target site duplications (TSDs). MITEs exhibit the structural features of class II transposons and are considered as truncated derivatives of autonomous class II transposons [7, 8]. Unlike autonomous DNA transposons, MITEs lack coding capacity and transpose through transposases provided in *trans* by their related autonomous elements [9].

Independent studies showed that MITEs could be mobilized by transposases from their related elements [10, 11]. Homology restricted to the TIRs and the sub-terminal sequences between MITEs and their related elements could be sufficient for cross-mobilization [12]. However, MITEs are present at a much higher copy number than autonomous DNA transposons, which mobilize them and from which they are derived, suggesting that MITEs are particularly successful in avoiding genome control. Yang et al. revealed that the MITE lacks a motif repressing transposition in the autonomous element and contains internal sequences that enhance transposition [13]. The amplification of autonomous DNA elements may be limited by a self-regulatory mechanism, while MITEs could achieve high transposition activity by scavenging transposases encoded by distantly related and self-restrained autonomous DNA elements [13]. The small size of MITEs may also help them to avoid silencing by host genomes [14]. Although MITEs are abundant in eukaryotic genomes, only very few MITEs have been found to be active in transposition likely because they are subject to purifying selection [14].

MITEs are grouped into different MITE families based on their size, structure, and sequence similarity between their TIRs or TSDs and these of autonomous partners. The structural homogeneity of MITE families suggests that they arose from amplification of a few progenitor copies. Major MITE superfamilies, such as *Tc1/Mariner*,

*PIF/Harbinger*, *hAT*, *Mutator*, and *CACTA*, have been described in plant genomes [9, 12, 13, 15]. Identification and classification of MITEs are mainly performed through searching sequences with TIR and TSD features. Bioinformatics programs, such as MITE-Hunter [16], MITE Digger [17], MITE tracker [18], detectMITE [19], and MAK [20], have been developed to identify MITEs from genome sequence databases.

MITEs are abundant in eukaryotic genomes and are thought to have a significant influence on the evolution of the host's genome structure. MITEs can mediate genomic rearrangements through insertion, excision, chromosome breakage, and ectopic recombination [21]. In addition, MITEs can affect gene function and regulation by gene transduction, duplication, exon shuffling, and insertion in gene regulatory regions [21, 22]. MITEs can change host gene expression by generating small RNAs, RNA-directed DNA methylation, and translational repression [23–25]. Moreover, MITEs also contribute to novel gene formation by providing start sites, poly(A) signals, splicing junctions, and TATA boxes [26, 27].

Pineapple (*A. comosus*) is the most economically important crop possessing crassulacean acid metabolism (CAM) and is a model for studying the evolution of CAM photosynthesis. The pineapple genome has one fewer ancient whole-genome duplication than grass genomes, providing an important reference for tracking evolutionary genomic changes and refining the evolutionary history of grass genomes [28]. In this study, we performed a genome-wide identification and characterization of MITEs in the pineapple genome for a better understanding of genome evolution.

## Results

### Identification and characterization of MITE families in the pineapple genome

We performed genome-wide identification of MITEs in the pineapple F153 reference genome using MITE-hunter. A total of 4659 representative MITE sequences were identified and they were further grouped into 243 MITE families (Additional file 1: Table S1). The consensus sequences of 243 MITE families (Additional file 2) were imported into RepeatMasker to scan all the associated MITE fragments in the pineapple genome. A total

of 212,351 MITE fragments were identified with a total length of 50,210,791 bp, accounting for approximately 13.14% of the pineapple genome. Among these MITE fragments, about 24.41% of them are intact (Table 1; Additional file 1: Table S1). The two largest MITE families, containing 53,014 elements and accounting for 29.39% of the total MITEs and 3.86% of the pineapple genome, were particularly analyzed in this study due to their special flanking sequences (Additional file 1: Table S1). Approximately 74% of them are flanked by TA dinucleotide microsatellites on both sides or one side, and 22 and 16.5% of them are flanked by GA or CT microsatellite, respectively (Additional file 3: Table S2). Therefore, we named these MITEs *A. comosus* microsatellite-associated MITEs (Ac-mMITEs). According to the phylogenetic analysis, Ac-mMITEs were divided into Ac-mMITE-1 and Ac-mMITE-2 (Fig. 1A, B; Additional file 4: Fig. S1), which shared sequence similarity in the terminal inverted repeat (TIR) regions (Additional file 5: Fig. S2; Additional file 6: Fig. S3), suggesting that these two Ac-mMITE families might be derived from a common ancestral or closely related autonomous elements.

In order to gain insight into the evolutionary dynamics of MITEs in the pineapple genome, we calculated Kimura distances (*K*-values) [29], which measure the degree of divergence between TE fragment and consensus. Low *K*-values suggest a relatively recent transposition event and activity. Our result showed that both Ac-mMITE-1 and Ac-mMITE-2 have a lower *K*-value than other MITEs, indicating that Ac-mMITEs have been created by recent transposition events (Fig. 1C). We further compared the sequence conservation and structural integrity of Ac-mMITEs with other MITEs in the pineapple genome. The Ac-mMITEs showed a higher level of sequence similarity and structural integrity than other MITEs (Fig. 1D, E). Taken together, our results imply that Ac-mMITEs have been generated by recent transposition bursts.

### Genomic distribution of Ac-mMITEs

More than 80% of Ac-mMITEs are flanked by dinucleotide microsatellites (Additional file 3: Table S2). We therefore investigated whether dinucleotide microsatellites are preferential target sites of Ac-mMITEs. We observed a strong correlation between the genomic distribution of Ac-mMITEs and (TA) n ( $R^2 = 0.6806$ ,

Fig. 2A, D), which suggests that (TA) n microsatellites were preferential target sites of Ac-mMITEs. We also observed a positive correlation between the genomic distribution of Ac-mMITEs and (GA) n and (TC) n microsatellites (Fig. 2E), but the  $R^2$  values are much lower than the one with (TA) n microsatellites. In addition, only 0.20 and 0.10% of Ac-mMITEs are flanked by (GA) n and (TC) n microsatellite on both sides, respectively (Additional file 3: Table S2). Most of the Ac-mMITEs associated with (GA) n and (TC) n microsatellites have (TA) n microsatellite on one side (Additional file 3: Table S2; Additional file 7: Fig. S4). Furthermore, the first and last two bases of the Ac-mMITE consensus sequences are mostly 'GA' and 'TC' (Additional file 5: Fig. S2). In consistent with this, most (GA) n microsatellites are located at the 5' end of Ac-mMITEs while most (TC) n microsatellites are located at the 3' end of Ac-mMITEs (Additional file 3: Table S2; Additional file 7: Fig. S4). All together suggest that (GA) n and (TC) n microsatellites might not be the preferential targets of Ac-mMITEs and the (GA) n and (TC) n microsatellites flanking Ac-mMITEs were likely generated by "DNA replication slippage" after the insertions of Ac-mMITEs.

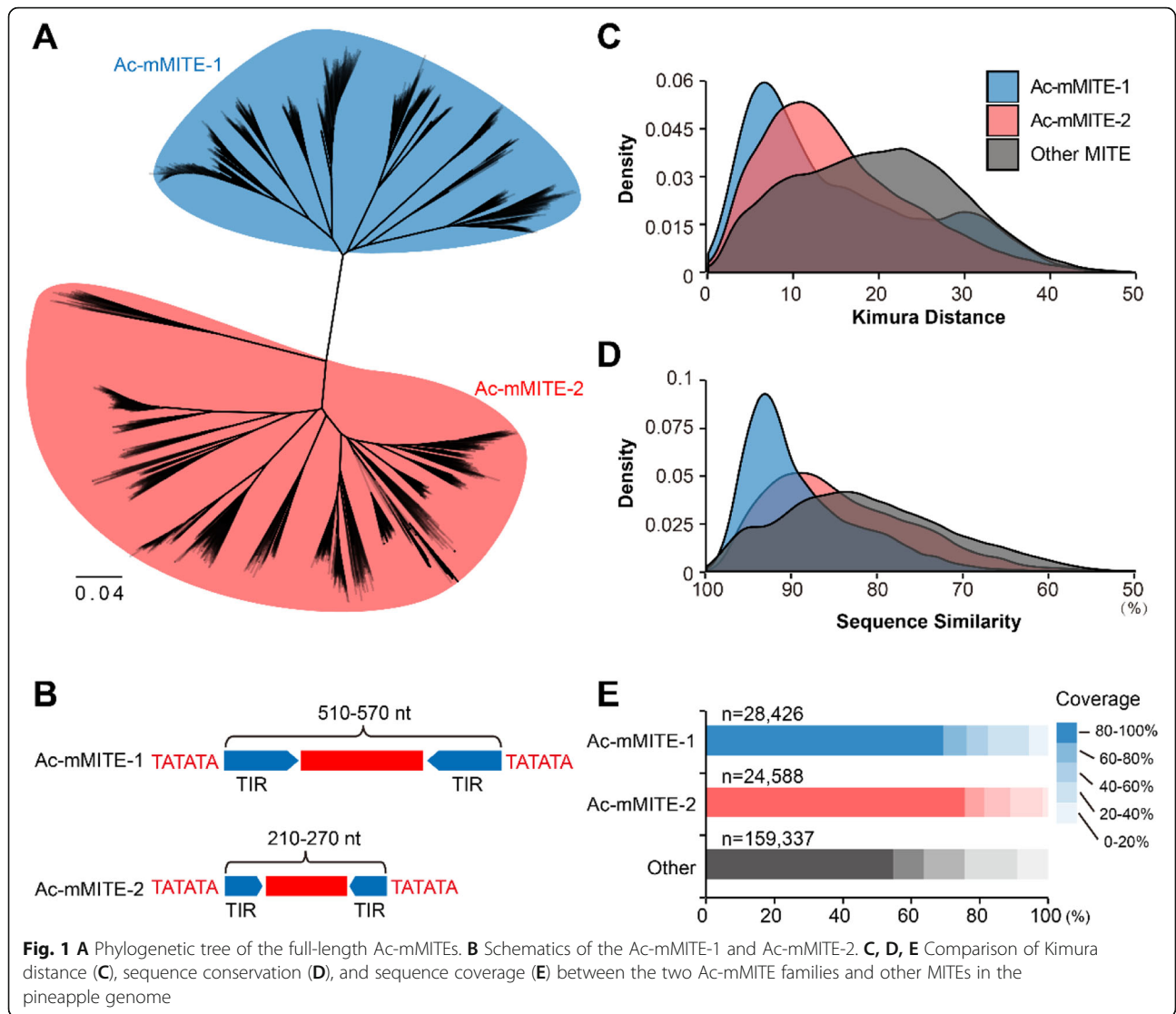
It has been reported that MITEs preferentially inserted into genic regions and significantly contributed to allelic diversity [30, 31]. We also tested whether there was a correlation of genomic distribution between Ac-mMITEs and genes. The Pearson Correlation Coefficient value calculated between Ac-mMITEs and genes is 0.0364 and much lower than the one calculated between other MITEs and genes ( $R^2 = 0.1682$ ) (Fig. 2A, B, C). In addition, the proportion of Ac-mMITEs that are located in intergenic regions is much higher than that of other MITEs, while the proportions of Ac-mMITEs that are located near or within genes are lower than that of other MITEs (Additional file 8: Table S3). Our results suggest that Ac-mMITEs prefer to target gene-sparse regions.

### Ac-mMITEs are related to the *Mutator* superfamily

Among the full-length Ac-mMITEs without dinucleotide microsatellites on both sides or one side, we discovered 1435 of them possess 9/10 bp TSDs. Given the feature of TSDs and TIRs of these Ac-mMITEs, we assumed that the Ac-mMITEs might be derived from *Mutator*-like transposable elements. We searched into pineapple genome to identify the corresponding autonomous elements

**Table 1** Summary of Ac-mMITEs and other MITEs in the pineapple genome

MITE families	Number of elements (intact elements)	Length (bp)	Total Length (bp)	% in genome	TSD	TIR
Ac-mMITE-1	28,426 (9889)	510–570	10,279,806	2.69%	(TA)n	140–180
Ac-mMITE-2	24,588 (12,105)	210–270	4,478,432	1.17%	(TA)n	40–70
Other MITEs	159,337 (29,851)	70–1300	35,452,553	9.28%	2-10 bp	8–400



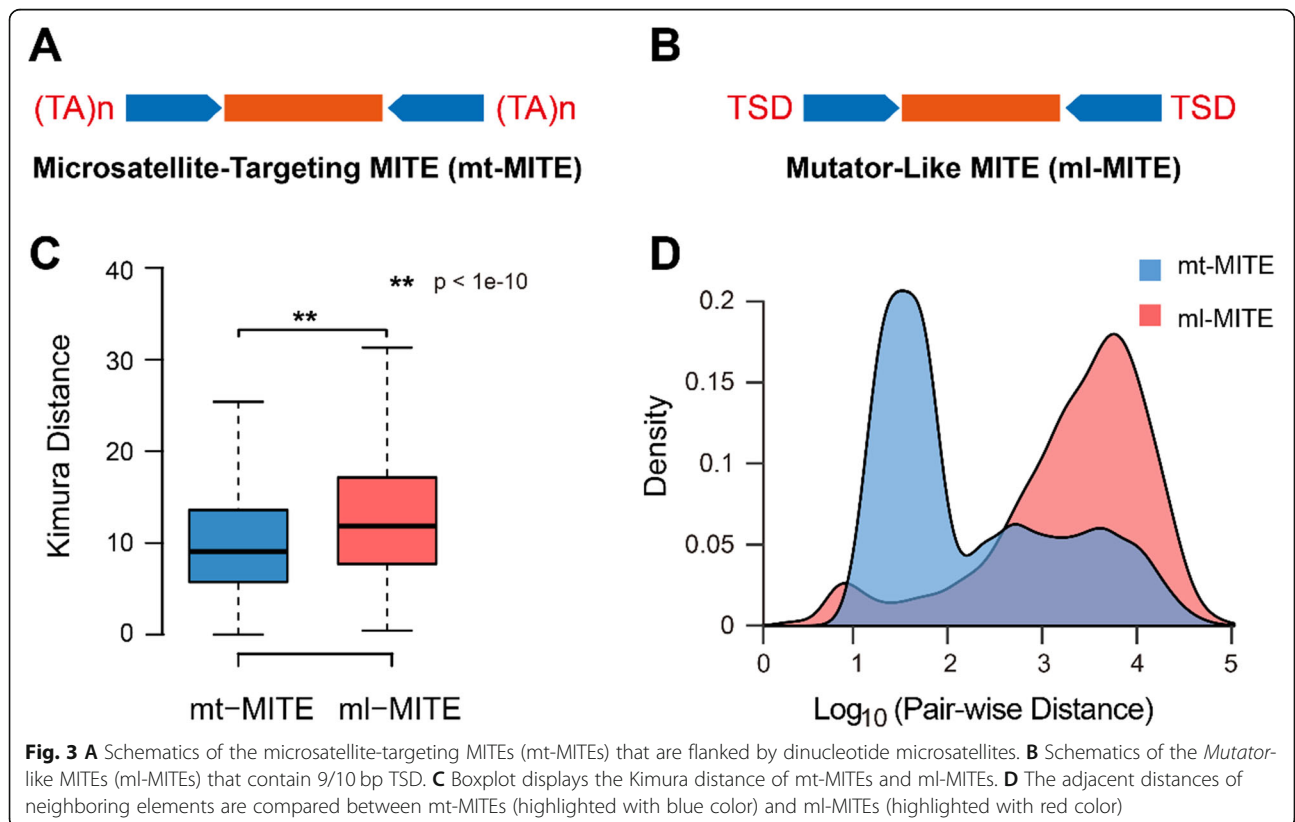
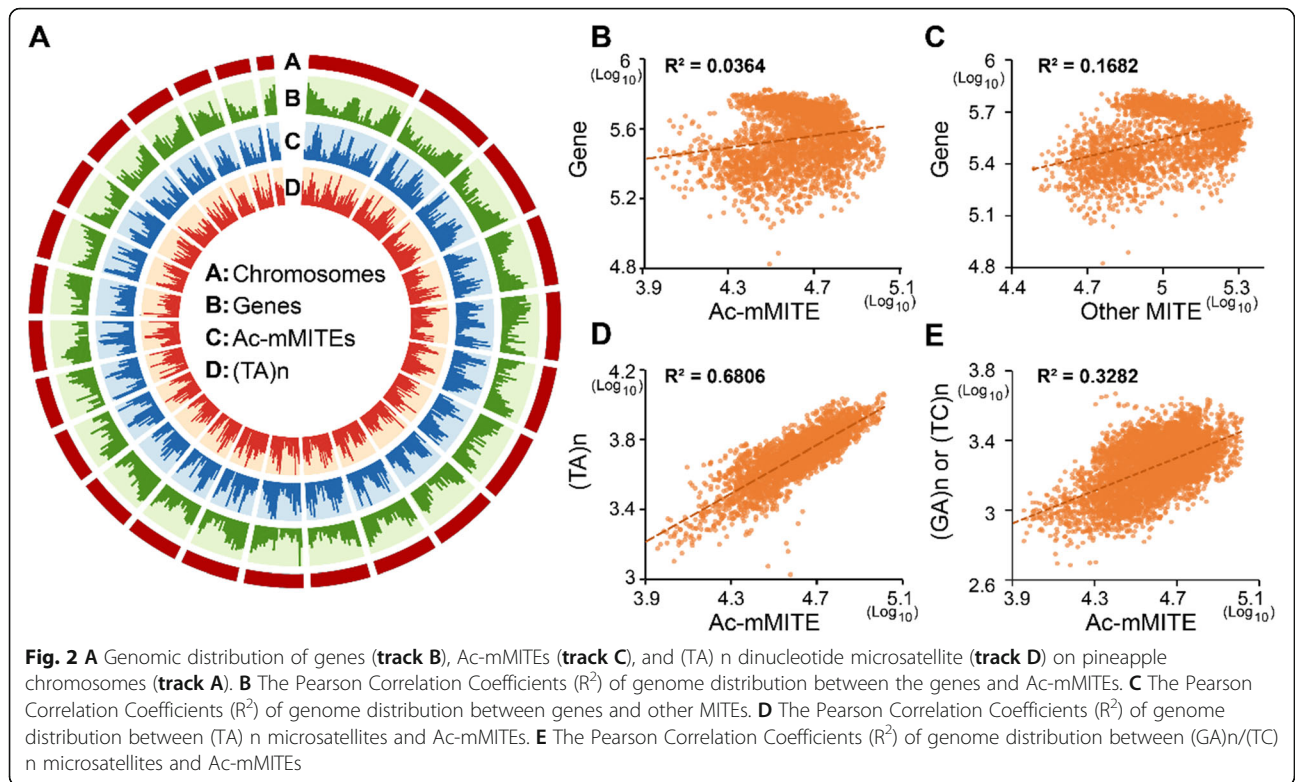
that provide the transposases required to facilitate transposition of Ac-mMITEs and no such elements were found, suggesting that the related autonomous elements might have largely mutated or degenerated.

To better understand the evolutionary history of Ac-mMITEs, we filtered Ac-mMITEs based on the features of flanking sequences and performed comparative analysis on the two distinct groups of Ac-mMITEs, microsatellite-targeting MITEs (mt-MITEs) that are flanked by dinucleotide microsatellites on both sides and *Mutator*-like MITEs (ml-MITEs) that contain 9/10 bp TSDs (Fig. 3A, B). The two Ac-mMITE families, Ac-mMITE-1 and Ac-mMITE-2, contained similar proportions of mt-MITEs and ml-MITEs. The copy number of mt-MITEs (15,361) is significantly larger than the ml-MITEs (1435), implying the higher activity of mt-MITEs over the ml-MITEs. Furthermore, the mt-MITEs showed a significantly lower *K*-value than the ml-MITEs (Fig. 3C), which suggests that the mt-

MITEs were generated by a more recent amplification burst compared to the ml-MITEs.

**mt-MITEs are frequently clustered via adjacent insertions**

We found a large number of intact Ac-mMITEs that are physically close to each other and linked via dinucleotide microsatellites in pineapple genome, indicating that Ac-mMITEs tend to form clusters by adjacent insertions. A cutoff of pair-wise distance of adjacent intact Ac-mMITEs within 100-bp was used to identify the Ac-mMITE clusters in pineapple genome (Additional file 9: Fig. S5). A total of 10,137 full-length Ac-mMITEs, accounting for 46.1% of the total full-length Ac-mMITEs, were screened out, which formed 4024 clusters. Interestingly, the Ac-mMITEs making up these clusters are non-nested and highly variable, indicating that these clusters were formed via multiple independent insertion events, not by tandem duplication. In addition, no

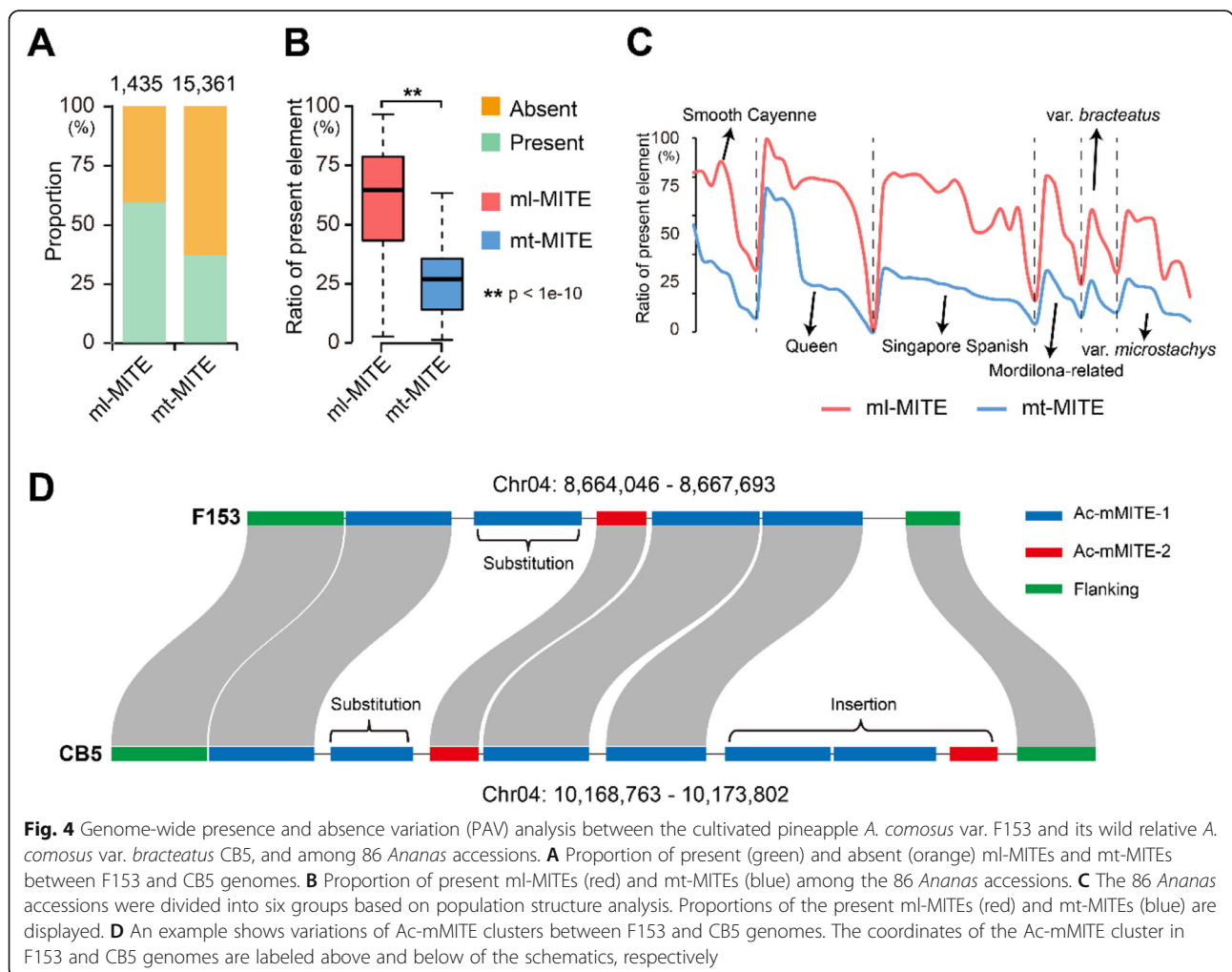


identical Ac-mMITE clusters were found in the pineapple genome, supporting that the entire Ac-mMITE cluster may not be capable of transposition. Furthermore, we observed that majority of the adjacent mt-MITEs are located within 100-bp while the ml-MITEs are sparsely distributed by a single unit in the pineapple genome, which is consistent with the discovery that the Ac-mMITE clusters are mostly composed of mt-MITEs (9573/10,137, Fig. 3D).

**Mt-MITEs are highly polymorphic between cultivated pineapples and their wild relatives**

To explore the transposition activity of Ac-mMITEs in the pineapple genome, we performed comparative analysis of Ac-mMITEs between the cultivated pineapple *A. comosus* var. F153 and its wild relative *A. comosus* var. *bracteatus* CB5. Ac-mMITEs account for 3.3% of the CB5 genome, which is at a similar level as in the F153 genome. The sequences of intact Ac-mMITEs in the F153 genome were used as reference to be compared with that in the

CB5 genome by performing genome-wide presence and absence variation (PAV) analysis. In total, we discovered 9089 intact Ac-mMITEs, including 5736 mt-MITEs and 851 ml-MITEs, that are present in the CB5 genome. Noticeably, we observed a lower proportion of mt-MITEs than ml-MITEs conserved between the two genomes (37.3% versus 59.3%, Fig. 4A), supporting that the mt-MITEs had experienced more frequent transposition compared to the ml-MITEs after the divergence of the two pineapple varieties from a common ancestor. We further performed PAV analysis using the sequences of the Ac-mMITE clusters in F153 genome as reference and the result revealed that 1123 and 605 clusters, accounting for 28 and 15% of the total clusters, were present and absent in the CB5 genome, respectively. Though the remaining clusters (2296/4024) can be found at the corresponding locations of the CB5 genome, these clusters have exhibited many variations between the two genomes (Fig. 4D). The high variability of these clusters between the two pineapple varieties could be ascribed to random transpositions



**Fig. 4** Genome-wide presence and absence variation (PAV) analysis between the cultivated pineapple *A. comosus* var. F153 and its wild relative *A. comosus* var. *bracteatus* CB5, and among 86 *Ananas* accessions. **A** Proportion of present (green) and absent (orange) ml-MITEs and mt-MITEs between F153 and CB5 genomes. **B** Proportion of present ml-MITEs (red) and mt-MITEs (blue) among the 86 *Ananas* accessions. **C** The 86 *Ananas* accessions were divided into six groups based on population structure analysis. Proportions of the present ml-MITEs (red) and mt-MITEs (blue) are displayed. **D** An example shows variations of Ac-mMITE clusters between F153 and CB5 genomes. The coordinates of the Ac-mMITE cluster in F153 and CB5 genomes are labeled above and below of the schematics, respectively

of mt-MITEs before or after formation of clusters, further demonstrating the recent high activity of mt-MITEs in the pineapple genome.

To further confirm the activity of mt-MITEs, we compared the degrees of insertion polymorphisms between mt-MITEs and ml-MITEs in 86 *Ananas* accessions. Consistent with our assumption, mt-MITEs showed a significantly lower proportion of present orthologous insertions than ml-MITEs (Fig. 4B, C). Based on the structural analysis of the *Ananas* population [32], we divided the *Ananas* accessions into six groups, including four representative groups in the var. *comosus* ('Queen', 'Smooth Cayenne', 'Singapore Spanish', and 'Mordilona-related'), one group of var. *bracteatus*, and one group of var. *microstachys*. The PAV patterns of Ac-mMITEs among the six groups match their origin and taxonomical relationships. The four groups within the var. *comosus* share a higher level of Ac-mMITEs than var. *bracteatus* and var. *microstachys*. Smooth Cayenne and Queen dispersed from the Guianas, while Singapore Spanish dispersed from the eastern coast of Brazil (south of Bahia) [33]. Smooth Cayenne and Queen groups share a relatively higher level of Ac-mMITEs than the other groups.

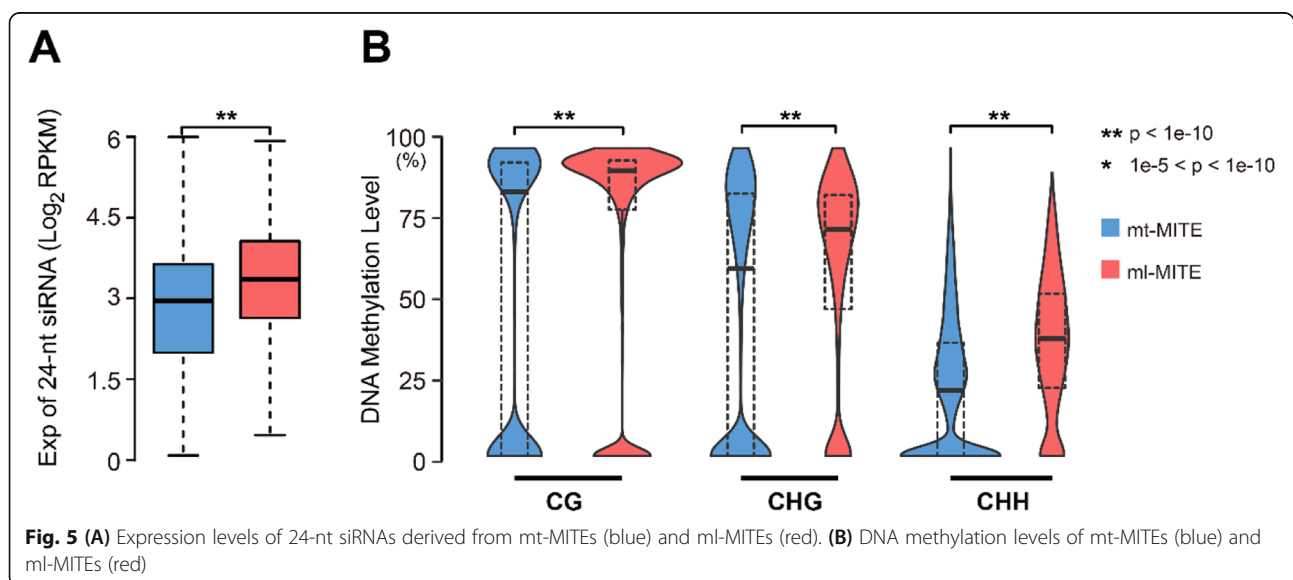
#### Differential epigenetic regulation of mt-MITEs and ml-MITEs in the pineapple genome

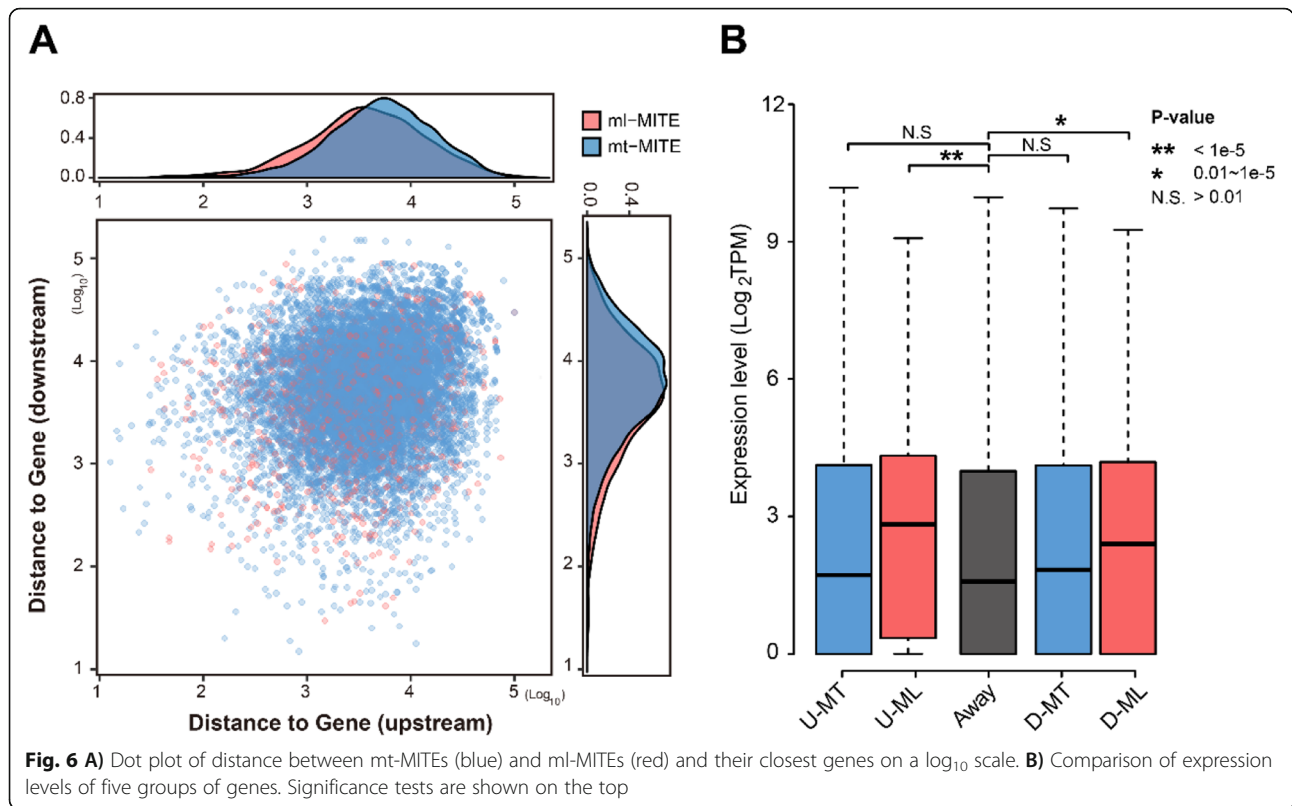
Due to the potential deleterious effects of TE insertions, host genomes usually silence TEs epigenetically through small-RNA-mediated DNA methylation to maintain genome integrity [34–36]. We employed the microRNA-seq data [37] (data are available at NCBI BioProject PRJNA311758) and the bisulfite sequencing data [38] (data are available at NCBI BioProject PRJNA493186) to investigate host response and epigenetic regulation of

Ac-mMITEs in the pineapple genome. The 24-nt siRNAs derived from mt-MITEs showed a significantly lower level than those from ml-MITEs (Student's T-test,  $p$ -value  $< 1e-10$ , Fig. 5A). In line with this, methylation levels of mt-MITEs were also significantly lower than ml-MITEs (Student's T-test,  $p$ -value  $< 1e-10$ , Fig. 5B). These results demonstrated the mt-MITEs were not regulated as strictly as the ml-MITEs, which possibly account for the successful amplification of mt-MITE in the pineapple genome.

Host genomes counteract TE activity by silencing them epigenetically, but methylation can spread beyond the TE sequence. It has been reported that MITEs have potential impact on gene expression [24, 30, 39]. In rice, genes with embedded or nearby MITEs showed lower levels of expression than the ones without MITE-gene interactions [40]. We discovered a longer distance between mt-MITEs and genes than that between ml-MITEs and genes (Fig. 6A), which is consistent with the lower proportion of mt-MITE assigned in 2-kb flanking regions of genes than that of the ml-MITEs (Table 2). These results suggested that the two kinds of Ac-mMITEs may have different effects on their proximal genes. To validate this assumption, we utilized the pineapple green leaf transcriptomic data (data are available at NCBI BioProject PRJNA493186) and compared expression levels of genes related to mt-MITEs and ml-MITEs separately.

In total, we identified 1688 and 1457 genes containing mt-MITE insertion in upstream (named as 'U-MT' group) and downstream (named as 'D-MT' group) regions, respectively, and 368 and 286 genes possessing ml-MITE insertion in upstream (named as 'U-ML' group) and downstream (named as 'D-ML' group) regions, respectively. A total of 17,135 genes that do not have Ac-mMITEs nearby (named as





‘AWAY’ group) were used as a reference group. No significant difference in expression levels was observed among U-MT, D-MT and AWAY groups, suggesting that mt-MITEs might have no or negligible impact on nearby gene expression (Student’s T-test, *p*-value > 0.01, Fig. 6B). However, the expression levels of both U-ML and D-ML groups were significantly higher than that of the other groups (Student’s T-test, *p*-value < 0.01, Fig. 6B).

**Discussion**

Transposable elements (TEs) constitute a significant fraction of plant genomes and play an important role in genome organization and evolution. Genome-wide identification and characterization of TEs provide essential information for understanding genome structure and evolution. Pineapple is largely vegetatively propagated. Sexual reproduction of pineapple is very rare in nature and is mainly restricted to breeding purpose. TEs might become a major source of genetic innovations in pineapple due to lack of recombination in asexually reproducing organisms [41]. MITEs are short DNA transposons. Although the

overall contribution of MITEs to the genome size is small, MITEs usually have high copy numbers [1, 4]. In addition, MITEs play important roles in gene expression and contribute considerable diversity [42].

We performed genome-wide identification and characterization of MITEs in the pineapple genome. The top two most abundant MITE families account for 29.39% of all MITEs and 3.86% of the pineapple genome. Interestingly, approximately 74% of these MITEs are flanked by (TA) n dinucleotide microsatellites, suggesting that they have insertion preference in (TA) n dinucleotide microsatellite regions. Furthermore, these MITEs frequently form non-nested clusters via adjacent insertions and the interval sequences between adjacent elements are almost pure (TA) n microsatellites, reinforcing the hypothesis that (TA) n dinucleotide microsatellite regions are the preferential target sites of Ac-mMITEs.

Mobilization of TEs can be highly mutagenic and cause genomic instability either by direct disruption of normal gene functions or by promoting ectopic homologous recombination, which can lead to harmful genome

**Table 2** The associations of mt-MITEs and ml-MITEs with genes

	Total	Intergenic	Upstream	Intron	Downstream
mt-MITE	15,361	9120 (59.4%)	2197 (14.3%)	2760 (18.0%)	1922 (12.5%)
ml-MITE	1435	672 (46.8%)	346 (24.1%)	265 (18.5%)	261 (18.2%)



rearrangements, deletions, and insertions [43, 44]. TEs with seriously deleterious effects on their host genomes will be mostly filtered out by natural selection. Host genomes have also evolved defense mechanisms to suppress TE activities, such as epigenetic silencing [45]. The interaction between TEs and defense mechanisms has led to an evolutionary arms race as well as self-control and targeting mechanism of TEs that mitigate the cost of their propagation on host fitness [46].

TEs are not evenly distributed across the genome and often exhibit various levels of preference of insertion [39, 47]. This may reflect the result of damage-limiting strategy adapted by TEs during co-evolution between TEs and their hosts. The evolutionary success of Ac-mMITEs may lie in their preferential insertion in (TA)<sub>n</sub> microsatellite regions. A strong bias of TE insertion towards (TA)<sub>n</sub> microsatellite repeats was also reported in rice [48], *M. truncatula* [49], guayule [50], and mammals [51]. Microsatellite repeats are predominantly non-coding sequences. TE insertion in these regions will have little or no impact on host genome and therefore may protect TEs from genome surveillance systems. In addition, (TA)<sub>n</sub> microsatellite regions are highly unstable [52], which may facilitate the integration and further transposition of TEs. Disruptions of these vulnerable regions by TE insertion may also increase the stability of these regions and provide potential benefits to host genomes.

The two Ac-mMITE families, Ac-mMITE-1 and Ac-mMITE-2, shared sequence similarity in the distal segments of the terminal inverted repeat (TIR) regions, suggesting that these two Ac-mMITE families were likely derived from a common or closely related autonomous elements. By analyzing the Ac-mMITEs without (TA)<sub>n</sub> microsatellite flanking sequences, we found that ml-MITEs were likely derived from *Mutator*-like DNA transposon. Ac-mMITEs showed a much higher proportion of intact elements and a lower *K*-value than other MITEs, suggesting that Ac-mMITEs were amplified through recent transposition bursts. mt-MITEs were much more abundant and showed a lower *K*-value than ml-MITEs, suggesting that their preferential insertion in (TA)<sub>n</sub> microsatellite regions occurred recently and insertion in (TA)<sub>n</sub> microsatellite regions might have promoted the amplification of mt-MITEs.

Polymorphic insertion analysis revealed highly polymorphic insertion sites of Ac-mMITEs among the 86 *Ananas* accessions. Surprisingly, highly polymorphic insertion sites of Ac-mMITEs were also observed in the close related accessions within the var. *comosus*. Highly divergent insertion of Ac-mMITEs might have resulted from their asexual reproduction and habitat isolation. The 86 *Ananas* accessions share a very low proportion of mt-MITEs. This suggests that mt-MITEs might have been mostly amplified after these accessions separated from a common ancestor and

transposition of mt-MITEs might have been ongoing since their divergence.

According to the general senescence patterns of TEs, young TEs are not yet silenced by the host genome and exhibit a low level or no CHH methylation, TEs at intermediate age are effectively silenced and usually show a high level of CHH methylation, and old TEs that are degenerated copies and unable to transpose are no longer silenced by the host genome. Our result showed that the CHH methylation level of mt-MITEs was significantly lower than that of ml-MITEs, providing a different line of evidence to support that mt-MITEs were mostly amplified recently.

TEs play important roles in the evolution of new genes and transcriptome diversity. TE insertions have potential impact on host gene expression through *cis*- or *trans*-regulatory activities [24, 30, 39, 53, 54]. Studies have implicated MITEs as negative transcription regulators of nearby genes [40]. However, MITEs may also upregulate gene expression by introducing regulatory motifs [39, 53, 54]. ml-MITEs showed a higher level of methylation than mt-MITEs. Surprisingly, genes nearby ml-MITEs showed a higher level of expression than the ones nearby mt-MITEs. Gene expression is controlled at multiple levels. Further studies are needed to address this issue. In general, TE insertions that significantly alter host gene expression patterns will be selected against. Therefore, TEs that cause minimal changes in host gene expression may help them escape the host genome control. Genes with and without mt-MITEs nearby showed similar levels of expression, which may also reflect the result of damage-limiting strategy adapted by mt-MITEs during co-evolution with their host.

## Conclusions

Ac-mMITEs are the most abundant MITEs in the pineapple genome and they were likely derived from *Mutator*-like DNA transposon. Preferential insertion in (TA)<sub>n</sub> microsatellite regions of Ac-mMITEs occurred recently and is likely the result of damage-limiting strategy adapted by Ac-mMITEs during co-evolution with their host. Insertion in (TA)<sub>n</sub> microsatellite regions might also have promoted the amplification of mt-MITEs. In addition, mt-MITEs showed no or negligible impact on nearby gene expression, which may help them escape genome control and lead to their amplification.

## Methods

### Identification and classification of MITEs in the pineapple genome

We used the MITE-Hunter program [16] to identify the MITEs in the genome assembly of the pineapple variety F153 [28] with default parameters. The putative MITEs were clustered into different families using VSEARCH

2.6.1 [55] with a parameter of 60% sequence similarity. The two largest MITE families, Ac-mMITE-1 and Ac-mMITE-2, represented by 45 consensus sequences generated by MITE-hunter, were used for further analysis. The flanking sequences of the Ac-mMITEs were manually trimmed using BioEdit [56]. The 45 consensus sequences representing the main subgroups of the Ac-mMITEs were used to scan the F153 genome assemblies using RepeatMasker 4.0.6 with a modified parameter of ‘-nolow -norna -no\_is -s -engine crossmatch’. Ac-mMITE fragments with a maximum missing of 10 bp from both terminals compared to the consensus sequences were considered full-length elements (Additional file 10: Fig. S65). The consensus sequences of Ac-mMITE-1 and Ac-mMITE-2 were used to predict the secondary structure of Ac-MITE using RNAstructure 6.0.1 [57, 58].

#### Estimation of divergence times

In order to estimate divergence times of Ac-mMITEs, we calculated pairwise Kimura distances [29] between Ac-mMITE copies and their corresponding consensus sequences using RepeatLandscape implemented in RepeatMasker. The transition and transversion rates were calculated on alignments generated by RepeatMasker and transformed to Kimura distance using the following equation:  $K = -1/2 \ln(1 - 2p - q) - 1/4 \ln(1 - 2q)$ , where  $q$  is the proportion of transversion sites and  $p$  is the proportion of transition sites. We also estimated sequence conservation by calculating similarities between Ac-mMITE sequences and their corresponding consensus sequences using EMBOSS Needle 6.6.0.0. Structural integrity of Ac-mMITEs was also assessed by calculating percent coverage of Ac-mMITEs aligned to their corresponding consensus sequences.

#### Construction of phylogenetic tree

To reduce the complexity of the dataset, we selected the top 20% of the full-length Ac-mMITEs with the highest sequence similarity to each of the 45 consensus sequences for constructing bootstrapped neighbor-joining trees using MEGA7 [59]. FigTree 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used for annotation and final graphic visualization of the phylogenetic tree.

#### Mining and characterization of dinucleotide microsatellites (TA) n, (CT) n, and (GA) n in the pineapple genome

We used the Tandem Repeat Finder 4.09 [60] to identify the dinucleotide microsatellites in the pineapple genome by modifying the parameters to ‘2 7 7 80 10 30 2’. Sliding window analysis (500-kb window size, 100-kb steps) was used to analyze the distributions of MITEs, genes,

and dinucleotide microsatellites (TA) n, (GA) n, and (TC) n across the pineapple chromosomes, and the results were visualized with Circos 0.69–6 [61].

#### Bisulfite sequencing (BS-seq) data analysis

Raw BS-seq reads of pineapple green leaf tip were downloaded from GEO under the accession number of GSE120401 [38]. BS-seq reads were mapped to the F153 reference genome using Bismark 0.20.0 [62] with default settings. The predicted methylation sites with less than 4 or more than 1000 supported reads were removed. The methylation level at each CpG site was obtained by estimating  $C/(C + T)$  ratio.

#### miRNA-seq and RNA-seq data analysis

Raw miRNA-seq reads of pineapple green leaf were downloaded from NCBI BioProject PRJNA311758 [37] (only the samples collected at 10:00 am were included in this analysis). We used Cutadapt 1.18 [63] to trim the raw miRNA-seq reads. The trimmed reads with length of 24-nt were then extracted and mapped to the pineapple reference genome using Bowtie 1.2.2 [64] with the modified parameters of ‘-v 0 -p 20 -m 2’. The reads that could be mapped to multiple locations were counted reciprocally, and the counted reads were normalized by Reads Per Kilobase per Million mapped reads (RPKM). Raw RNA-seq reads of pineapple green leaf tip were downloaded from GEO [37] (accession number: GSE120401). We used Bowtie2 2.3.4.1 [65] and RSEM 1.2.29 [66] to map reads and quantify transcripts with default settings. mRNA abundance was then normalized by ‘Transcripts Per Million’ (TPM).

#### Ac-mMITE insertion polymorphism analysis

To analyze the presence/absence variations (PAVs) of Ac-mMITEs between the pineapple F153 and CB5 reference genomes, the full-length Ac-mMITEs with 200 bp flanking sequences were extracted from the F153 genome, which was further used as a seed to search into the CB5 genome using NCBI-blastn with a modified parameter of ‘-xdrop\_gap 1000 -culling\_limit 1 -evalue 1e-100’. An Ac-mMITE was considered ‘present’ in the CB5 genome if the Ac-mMITE with 200 bp flanking sequences can be found at the corresponding position in the CB5 genome with at least 90% sequence similarity. Otherwise, it was marked as ‘absence’.

We further surveyed polymorphisms of the Ac-mMITEs among *Ananas* population of 86 resequencing accessions. The raw reads of 86 *Ananas* NGS data were downloaded from the NCBI BioProject database under the accession number PRJNA389669 [32]. The clean reads were mapped to the F153 genome using Bowtie2 with default parameters. An Ac-mMITE was marked as ‘present’ when there was at least one pair-end reads

covering the entire sequence of the Ac-mMITE and 50 bp flanking regions.

To investigate the polymorphisms of Ac-mMITE clusters between the F153 and CB5 genomes, the entire clusters with 200 bp flanking sequences were extracted from the F153 genome and used to run comparative analysis in the CB5 genome. The clusters were considered as 'present' in the CB5 genome when: **i)** the number and order of elements are identical. **ii)** the orientation and classification (Ac-mMITE-1 or Ac-mMITE-2) of elements in the cluster are identical; **iii)** each pair of elements display at least 90% of sequence similarity; **iv)** the flanking sequences of the two homologous clusters must have at least 90% of sequence similarity. Otherwise, they were considered 'absent' in the CB5 genome. It was defined as a deletion or insertion event when one/few elements absent or one/few additional elements present in a cluster at the corresponding location in the CB5 genome. It was defined as a substitution event when corresponding elements share no or very low sequence similarity or they belong to different Ac-mMITE families.

#### Abbreviations

MITEs: Miniature inverted-repeat transposable elements; Ac-mMITEs: *A. comosus* microsatellite-associated MITEs; TEs: Transposable elements; TIRs: Terminal inverted repeats; TSDs: Target site duplications; CAM: Crassulacean acid metabolism; mt-MITEs: Microsatellite-targeting MITEs; ml-MITEs: *Mutator*-like MITEs; PAV: Presence and absence variation

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-021-03194-0>.

**Additional file 1: Table S1.** Summary of MITE families in the pineapple F153 genome.

**Additional file 2:** Consensus sequences of all MITE families in the pineapple genome.

**Additional file 3: Table S2.** Summary of flanking sequences of Ac-mMITEs.

**Additional file 4: Figure S1.** Secondary structure analysis of Ac-mMITEs. The consensus sequences of Ac-mMITE-1 (**A**) and Ac-mMITE-2 (**B**) were used to predict secondary structure. The red arrows mark the end of terminal inverted repeats (TIRs).

**Additional file 5: Figure S2.** Alignment of TIR regions of 45 Ac-mMITE-1 and Ac-mMITE-2 consensus sequences. The TIR sequences of the two Ac-mMITE families share sequence similarity at the first 55 bases of 5' TIR and last 55 bp of the 3' TIR regions.

**Additional file 6: Figure S3.** Ac-mMITEs show a much higher level of sequence similarity at 5' and 3' TIR regions than the middle region. Sequence similarities were calculated based on the 45 consensus sequences representing the main subgroups of Ac-mMITEs using sliding windows of 30-bp windows and 5-bp steps.

**Additional file 7: Figure S4.** (**A**) Three possible insertions that resulted in Ac-mMITEs flanked by (TA) n (**i**), (GA) n (**ii**) and (CT) n (**iii**). (**B**) Most of the (GA) n are located at 5' end (**i**) and the (CT) n are located at 3' end (**iii**) of the Ac-mMITEs.

**Additional file 8: Table S3.** Association of Ac-mMITEs and other MITEs with genes.

**Additional file 9: Figure S5.** The number of Ac-mMITEs with the adjacent distance increasing every 10 bp is shown.

**Additional file 10: Figure S6.** Variations in the two ends of TIRs of the 21,994 intact Ac-mMITEs.

#### Acknowledgements

We thank Ratnesh Singh for helpful discussions. The open access publishing fees for this article have been covered in part by the Texas A&M University Open Access to Knowledge Fund (OAKFund), supported by the University Libraries.

#### Authors' contributions

Q.Y. lead this research project and coordinated all the research activities. L.L. and A.S. designed and performed the research and data analysis. L.L., Q.Y., and A.S. wrote the manuscript. All the authors reviewed and approved the final manuscript.

#### Funding

This work was supported by the United States Department of Agriculture National Institute of Food and Agriculture Hatch Project TEX0-2-9374 and Multi State Hatch Project NC1200 to Q.Y. The funding body had no role in the design of this study, collection, analysis, and interpretation of the data, and in the writing of the manuscript.

#### Availability of data and materials

All the datasets used in this study are publicly available at the National Center of Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>). The pineapple F153 reference genome was published by Ming et al. [25] and the sequence data is available in NCBI GenBank assembly under project PRJNA305080. Resequencing of the 86 *Ananas* accessions was published by Chen et al. [29] and the sequence data is available in NCBI BioProject database under the accession number PRJNA389669. The bisulfite sequencing of pineapple green leaf was published by Shi et al. [35] and the sequence data is available in NCBI GEO database under the accession number GSE120401. The miRNA-seq of pineapple green leaf was published by Wai et al. [34] and the sequence data is available in NCBI BioProject database under the accession number PRJNA311758.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Texas A&M AgriLife Research Center at Dallas, Texas A&M University System, Dallas, TX 75252, USA. <sup>2</sup>College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

Received: 21 May 2021 Accepted: 9 August 2021

Published online: 18 September 2021

#### References

- Bureau TE, Wessler SR. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell*. 1994;6(6):907–16. <https://doi.org/10.1105/tpc.6.6.907>.
- Braquart C, Royer V, Bouhin H. DEC: a new miniature inverted-repeat transposable element from the genome of the beetle *Tenebrio molitor*. *Insect Mol Biol*. 1999;8(4):571–4. <https://doi.org/10.1046/j.1365-2583.1999.00144.x>.
- Tu Z. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A*. 1997;94(14):7475–80. <https://doi.org/10.1073/pnas.94.14.7475>.

4. Smit AF, Riggs AD. Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci U S A*. 1996;93(4):1443–8. <https://doi.org/10.1073/pnas.93.4.1443>.
5. Yeadon PJ, Catcheside DE. Guest: a 98 bp inverted repeat transposable element in *Neurospora crassa*. *Mol Gen Genet MGG*. 1995;247(1):105–9. <https://doi.org/10.1007/BF00425826>.
6. Sun C, Feschotte C, Wu Z, Mueller RL. DNA transposons have colonized the genome of the giant virus *Pandoravirus salinus*. *BMC Biol*. 2015;13(1):1–12.
7. Oosumi T, Garlick B, Belknap WR. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J Mol Evol*. 1996;43(1):11–8. <https://doi.org/10.1007/BF02352294>.
8. Feschotte C, Mouches C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Mol Biol Evol*. 2000;17(5):730–7. <https://doi.org/10.1093/oxfordjournals.molbev.a026351>.
9. Zhang X, Feschotte C, Zhang Q, Jiang N, Eggleston WB, Wessler SR. P instability factor: an active maize transposon system associated with the amplification of tourist-like MITEs and a new superfamily of transposases. *Proc Natl Acad Sci*. 2001;98(22):12572–7. <https://doi.org/10.1073/pnas.211442198>.
10. Dufresne M, Hua-Van A, Abd el Wahab H, M'Barek SB, Vasnier C, Teyssier L, Kema GH, Daboussi M-J. Transposition of a fungal miniature inverted-repeat transposable element through the action of a Tc1-like transposase. *Genetics*. 2007;175(1):441–52. <https://doi.org/10.1534/genetics.106.064360>.
11. Miskey C, Papp B, Mátés L, Sinzelle L, Keller H, Izsvák Z, et al. The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Mol Cell Biol*. 2007;27(12):4589–600. <https://doi.org/10.1128/MCB.02027-06>.
12. Feschotte C, Swamy L, Wessler SR. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics*. 2003;163(2):747–58. <https://doi.org/10.1093/genetics/163.2.747>.
13. Yang G, Nagel DH, Feschotte C, Hancock CN, Wessler SR. Tuned for transposition: molecular determinants underlying the hyperactivity of a stowaway MITE. *Science*. 2009;325(5946):1391–4. <https://doi.org/10.1126/science.1175688>.
14. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res*. 2009;19(8):1419–28. <https://doi.org/10.1101/gr.091678.109>.
15. Moreno-Vázquez S, Ning J, Meyers BC. 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(6):869–86. <https://doi.org/10.1007/s11103-005-8271-8>.
16. Han Y, Wessler SR. MITE-hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res*. 2010;38(22):e199. <https://doi.org/10.1093/nar/gkq862>.
17. Yang G. MITE digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements. *BMC bioinformatics*. 2013;14(1):186. <https://doi.org/10.1186/1471-2105-14-186>.
18. Crescente JM, Zavallo D, Helguera M, Vanzetti LS. MITE tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC bioinformatics*. 2018;19(1):348. <https://doi.org/10.1186/s12859-018-2376-y>.
19. Ye C, Ji G, Liang C. detectMITE: a novel approach to detect miniature inverted repeat transposable elements in genomes. *Sci Rep*. 2016;6(1):19688. <https://doi.org/10.1038/srep19688>.
20. Yang G, Hall TC. MAK, a computational tool kit for automated MITE analysis. *Nucleic Acids Res*. 2003;31(13):3659–65. <https://doi.org/10.1093/nar/gkg531>.
21. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*. 2007;41(1):331–68. <https://doi.org/10.1146/annurev.genet.40.110405.090448>.
22. 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(11):1160–3. <https://doi.org/10.1038/ng.942>.
23. Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, et al. 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(1):42–56. <https://doi.org/10.1101/gr.078196.108>.
24. Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, et al. Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc Natl Acad Sci U S A*. 2014;111(10):3877–82. <https://doi.org/10.1073/pnas.1318131111>.
25. Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, et al. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat Commun*. 2015;6(1):1–13.
26. 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(3):1005–17. <https://doi.org/10.1093/molbev/msr282>.
27. 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 Genetic Syst*. 2008;83(4):321–9. <https://doi.org/10.1266/ggs.83.321>.
28. Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, et al. The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet*. 2015;47(12):1435–42. <https://doi.org/10.1038/ng.3435>.
29. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111–20. <https://doi.org/10.1007/BF01731581>.
30. Feschotte C, Jiang N, Wessler SR. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet*. 2002;3(5):329–41. <https://doi.org/10.1038/nrg793>.
31. Zhang Q, Arbuckle J, Wessler SR. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family heartbreaker into genic regions of maize. *Proc Natl Acad Sci U S A*. 2000;97(3):1160–5. <https://doi.org/10.1073/pnas.97.3.1160>.
32. Chen LY, VanBuren R, Paris M, Zhou H, Zhang X, Wai CM, et al. The bracteatus pineapple genome and domestication of clonally propagated crops. *Nat Genet*. 2019;51(10):1549–58. <https://doi.org/10.1038/s41588-019-0506-8>.
33. Ming R. *Genetics and genomics of pineapple*: springer; 2018. <https://doi.org/10.1007/978-3-030-00614-3>.
34. Rigal M, Mathieu O. A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta*. 2011;1809(8):452–8. <https://doi.org/10.1016/j.bbagr.2011.04.001>.
35. Matzke MA, Moshier RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*. 2014;15(6):394–408. <https://doi.org/10.1038/nrg3683>.
36. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010;11(3):204–20. <https://doi.org/10.1038/nrg2719>.
37. Wai CM, VanBuren R, Zhang J, Huang L, Miao W, Edger PP, et al. Temporal and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple. *Plant J*. 2017;92(1):19–30. <https://doi.org/10.1111/tpj.13630>.
38. Shi Y, Zhang X, Chang X, Yan M, Zhao H, Qin Y, et al. Integrated analysis of DNA methylome and transcriptome reveals epigenetic regulation of CAM photosynthesis in pineapple. *BMC Plant Biol*. 2021;21(1):1–14.
39. Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*. 2009;461(7267):1130–4. <https://doi.org/10.1038/nature08479>.
40. Chen J, Lu C, Zhang Y, Kuang H. Miniature inverted-repeat transposable elements (MITEs) in rice were originated and amplified predominantly after the divergence of *Oryza* and *Brachypodium* and contributed considerable diversity to the species. *Mob Genet Elem*. 2012;2(3):127–32. <https://doi.org/10.4161/mge.20773>.
41. Ferreira de Carvalho J, de Jager V, van Gurp TP, Wagemaker NC, Verhoeven KJ. Recent and dynamic transposable elements contribute to genomic divergence under asexuality. *BMC Genomics*. 2016;17(1):884.
42. Wessler SR, Bureau TE, White SE. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev*. 1995;5(6):814–21. [https://doi.org/10.1016/0959-437X\(95\)80016-X](https://doi.org/10.1016/0959-437X(95)80016-X).
43. Hedges DJ, Deininger PL. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res*. 2007;616(1–2):46–59. <https://doi.org/10.1016/j.mrfmmm.2006.11.021>.
44. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol*. 2014;65(1):505–30. <https://doi.org/10.1146/annurev-arplant-050213-035811>.

45. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8(4):272–85. <https://doi.org/10.1038/nrg2072>.
46. Cosby RL, Chang NC, Feschotte C. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* 2019;33(17–18):1098–116. <https://doi.org/10.1101/gad.327312.119>.
47. Dietrich CR, Cui F, Packila ML, Li J, Ashlock DA, Nikolau BJ, et al. Maize mu transposons are targeted to the 5' untranslated region of the gl8 gene and sequences flanking mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics.* 2002;160(2):697–716. <https://doi.org/10.1093/genetics/160.2.697>.
48. Akagi H, Yokozeki Y, Inagaki A, Mori K, Fujimura T. Micron, a microsatellite-targeting transposable element in the rice genome. *Mol Gen Genomics.* 2001;266(3):471–80. <https://doi.org/10.1007/s004380100563>.
49. 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(4):433–40. <https://doi.org/10.1007/s10709-015-9842-5>.
50. Valdes Franco JA, Wang Y, Huo N, Ponciano G, Colvin HA, McMahan CM, et al. Modular assembly of transposable element arrays by microsatellite targeting in the guayule and rice genomes. *BMC Genomics.* 2018;19(1):271. <https://doi.org/10.1186/s12864-018-4653-6>.
51. Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA. High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol.* 2005;25(6):2085–94. <https://doi.org/10.1128/MCB.25.6.2085-2094.2005>.
52. van Wietmarschen N, Sridharan S, Nathan WJ, Tubbs A, Chan EM, Callen E, et al. Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature.* 2020;586(7828):292–8. <https://doi.org/10.1038/s41586-020-2769-8>.
53. Yang G, Lee YH, Jiang Y, Shi X, Kertbundit S, Hall TC. A two-edged role for the transposable element kiddo in the rice ubiquitin2 promoter. *Plant Cell.* 2005;17(5):1559–68. <https://doi.org/10.1105/tpc.104.030528>.
54. Li J, Liu C, Zhao A, Yu M, Liu X, Chen X, et al. A MITE insertion in the promoter region of Anthocyanidin synthase from *Morus alba* L. *Plant Mol Biol Report.* 2018;36(2):188–94. <https://doi.org/10.1007/s11105-018-1069-z>.
55. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 2016;4:e2584. <https://doi.org/10.7717/peerj.2584>.
56. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41: 95–8.
57. Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.* 2013;41(Web Server issue):W471–4.
58. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics.* 2010;11(1):129. <https://doi.org/10.1186/1471-2105-11-129>.
59. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4. <https://doi.org/10.1093/molbev/msw054>.
60. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>.
61. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–45. <https://doi.org/10.1101/gr.092759.109>.
62. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571–2. <https://doi.org/10.1093/bioinformatics/btr167>.
63. Martin MG. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10–2. <https://doi.org/10.14806/ej.17.1.200>.
64. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics.* 2010; Chapter 11:Unit 11 17.
65. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
66. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12(1):323. <https://doi.org/10.1186/1471-2105-12-323>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

