

Genome-wide analyses of tandem repeats and transposable elements in patchouli

Linqiu Liu, Junjun Li, Jiawei Wen and Yang He*

State Key Laboratory of Southwestern Chinese Medicine Resources, College of Medical Technology,
Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan 611137, China

(Received 14 August 2020, accepted 3 January 2021; J-STAGE Advance published date: 22 April 2021)

Patchouli, *Pogostemon cablin* (Blanco) Benth., is a traditional Chinese medicinal plant from the order Lamiales. It is considered a valuable herb due to its essential oil content and range of therapeutic effects. This study aimed to explore the evolutionary history of repetitive sequences in the patchouli genome by analyzing tandem repeats and transposable elements (TEs). We first retrieved genomic data for patchouli and four other Lamiales species from the GenBank database. Next, the content of tandem repeats with different period sizes was identified. Long terminal repeats (LTRs) were then identified with LTR_STRUC. Finally, the evolutionary landscape of TEs was explored using an in-house PERL program. The analysis of repetitive sequences revealed that tandem repeats constitute a higher proportion of the patchouli genome compared to the four other species. Analyses of TE families showed that most of the repetitive sequences in the patchouli genome are TEs, and that recently inserted TEs make up a comparatively larger proportion than older ones. Our analyses of LTR retrotransposons in their host genome indicated the existence of ancient LTR retrotransposon expansion, and the escape of these elements from natural selection revealed their ages. Our identification and analyses of repetitive sequences should provide new insights for further investigation of patchouli evolution.

Key words: evolutionary history, patchouli, repetitive sequence, tandem repeats, transposable elements

INTRODUCTION

Patchouli, *Pogostemon cablin* (Blanco) Benth., is a plant belonging to the genus *Pogostemon* of the Lamiales. Originating from Southeast Asia, its introduction into China can be dated back to the Liang Dynasty (Wu et al., 2007). For many years, patchouli has been widely used in medicinal materials, with great prospects in medicinal production and applications. Its antiemetic and heat-releasing actions (He et al., 2016) are commonly used in traditional Chinese medicine. As an increasing number of technologies developed, modern studies

have observed that patchouli and its essential oil have a large number of applications in medical fields, including antibacterial, antitumor, antioxidant and insecticidal effects, and also in chemical fields, as important materials (Chinese Pharmacopoeia Commission, 2010).

Repetitive sequences are ubiquitous in many species and often account for a major proportion of eukaryotic genomes (Britten and Kohne, 1968). Repeats in a genome can be divided into tandem repeats and interspersed repeats. Tandem repeat sequences connect end to end to form a repetitive sequence with a relatively constant short sequence as a repeating unit, and play roles in both human diseases and genome evolution (Hatters and Hannan, 2013). Transposable elements (TEs), one subtype of interspersed repeats, are mobile genetic elements. Widespread and abundant in almost all genomes of eukaryotic species, TEs are greatly influential in the evolution and structural organization of genes and genomes (Feschotte et al., 2002; Bennetzen, 2005; Biémont and Vieira, 2006; Feschotte, 2008; Bucher et al., 2012). TEs can be further categorized into two classes (Finnegan, 1989). Class I TEs use an RNA intermediate

Edited by Hideki Innan

* Corresponding author. heyang@cdutcm.edu.cn

DOI: <https://doi.org/10.1266/ggs.20-00044>



Copyright: ©2021 The Author(s). This is an open access article distributed under the terms of the Creative Commons BY 4.0 International (Attribution) License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits the unrestricted distribution, reproduction and use of the article provided the original source and authors are credited.

to transpose through a copy and paste mechanism, while class II TEs transpose by excising from one site and moving to another one by a cut and paste mechanism. Class I TEs are also called retrotransposons and include LTR retrotransposons, short interspersed nuclear elements and long interspersed nuclear elements; LTR retrotransposons possess LTRs (long terminal repeats) at both sides of the element. Moreover, previous studies have recognized the critical role played by LTR retrotransposons in the evolutionary history of numerous species (Beulé et al., 2015; Yin et al., 2015). The *Gypsy* and *Copia* families, as members of the LTR retrotransposons, are highly predominant in the genomes of flowering plants (Kumar and Bennetzen, 1999; Wicker et al., 2007). Furthermore, TE activities such as insertion and deletion may affect the regulation of adjacent genes, which may in turn give rise to phenotypic variation.

As previous research has demonstrated, with the development of high-throughput DNA sequencing techniques, a considerable number of bioinformatic approaches have been established to identify abundant repetitive sequences. For instance, repeat-search programs found a large number of repetitive DNA motifs (Biscotti et al., 2015). Currently, studies involving key genes and the genome of patchouli are expanding (An et al., 2019). However, these studies have focused on key genes rather than patchouli repeats. To fill in the existing knowledge gap concerning repeats in the patchouli genome, the present study aimed to analyze repetitive sequences in this genome, setting it apart from previous research. Previous research found the contribution of repeats to the field of DNA sequence analysis, as well as developing bioinformatics technology to analyze LTR retrotransposons (Kidwell and Lisch, 2001). With these tools, this paper addresses tandem repeats and TEs rather than only LTR retrotransposons.

To understand the relationship between the repetitive sequences in patchouli and the evolutionary history, *de novo* repeat annotation was collected from existing data, and the tandem repeats and TEs from the patchouli genome were compared with those of four other species. The patchouli genome with its repetitive sequences offers a reference to deduce information about the most recent common ancestor of all extant Lamiaceae. Evolutionary analysis of patchouli is thus important to elucidate the evolution of Lamiales at the genome level (Kumar and Bennetzen, 1999; Wicker et al., 2007). Collectively, the principal issue addressed in this paper was to define a unique evolutionary history of repetitive sequences in the patchouli genome. Exploration of patchouli evolution may also benefit analysis of its medicinal traits. It is therefore important to shed new light on the repeats and on their content, evolution and unknown function in plant metabolism.

MATERIALS AND METHODS

Data resources for genomes and annotation We chose four species to investigate in this study as they are phylogenetically closely related to patchouli (He et al., 2018). The unmasked and raw whole genomic sequences of patchouli were collected from the GenBank database (<https://www.ncbi.nlm.nih.gov/>) with the accession number QKXD000000000. Gene and repeat annotation information of the patchouli genome was downloaded from the website (<https://doi.org/10.6084/m9.figshare.c.4100495>) that we completed in our previous research (He et al., 2018). Genome data for other species were acquired from existing resources, and included *Salvia miltiorrhiza* (<ftp://202.203.187.112>, accession number: PRJNA287594), *Utricularia gibba* (<http://genomeevolution.org>, accession number: NEEC000000000), *Mimulus guttatus* (<http://phytozome.jgi.doe.gov/>, accession number: APLE000000000) and *Sesamum indicum* (<http://www.ocri-genomics.org>, accession number: APMJ000000000) (Hellsten et al., 2013; Ibarra-Laclette et al., 2013; Wang et al., 2014; Zhang et al., 2015; Lan et al., 2017).

Identification of tandem repeats Each genome was annotated by the tandem repeats finder (TRF) using default parameters (Benson, 1999) to compare the tandem repeats in patchouli and its relatives. Merged fragments from the read-through library with an insert size of 250 bp were also annotated by the TRF. This two-part identification process includes detection and analysis of components. First, we used a series of criteria based on statistics to detect those participant tandem repeats with k-tuple matches mentioned in previous studies (Benson, 1999). Next, we aimed to align each set of candidate repeats and to make full sense of proper parameters in the process of alignment. Finally, the relationship between the percentage of the genome comprised of these tandem repeats and the period size was described. The period size is defined as the optimum matching distance of these corresponding elements during alignment. Using the same approach, we identified tandem repeats of *Sa. miltiorrhiza*, *U. gibba*, *M. guttatus* and *Se. indicum*.

TE expansion history analysis An in-house PERL script was written to parse the result (.out file) generated by RepeatMasker (<http://www.repeatmasker.org>) to determine the divergence between copies of the TEs identified in the genome and the consensus sequence in the library. Sequence divergence is usually denoted as K, which can be obtained by sequence pair comparison and corrected by a sequence evolution model (divergence K can be calculated by RepeatMasker). The percentage of each TE family in each divergence window, with a window size of 0.01 ranging from 0.0 to 0.60, was calculated. Transposable element families that constituted

less than 0.1% of the genome were excluded from further analysis. An excessive accumulation of the TE at a certain evolutionary time point indicated a potential TE expansion in the host genome. To detect expansion of a TE, we used the percent divergence as a proxy for the age of the TE. Therefore, if the divergence between the consensus and the different TE paralogs is small, this is indicative of recent TE activity (transposition). Conversely, greater divergence is indicative of the absence of a recent TE burst. In other words, the master copy or (copies) that is responsible for the burst may be an ancient TE(s).

Analysis of LTR retrotransposons LTR_STRUC (McCarthy and McDonald, 2003) is a structure-based tool that can both scan nucleotide sequence files for LTR retrotransposons and also analyze any resulting hits. These elements were identified and automatically analyzed through the tool's features based on the process of searching for and aligning structural features of LTR retrotransposons in the genome database. To identify the LTR retrotransposons, full-length LTR retrotransposons were retrieved from LTR_STRUC to clarify their architecture, and the time of insertion was calculated as described previously (Hu et al., 2011). In detail, insertion times were estimated on the basis of the divergence between both LTR sequences of the same element. Intact LTR retrotransposons were then identified using LTR_STRUC with default parameters, and LTR pairs were aligned using MUSCLE (Edgar, 2004). Using the Kimura two-parameter model as exploited in the EMBOSS package with the distmat program (<http://emboss.sourceforge.net/>), the distance K between the pairs was calculated with the formula $T = K/2r$, where T is the insertion time and r the nucleotide substitution rate (Du et al., 2010). The mutation rate was assumed to be 8.1×10^{-9} substitutions per site per year. The LTR retrotransposons of *P. cablin*, *Sa. miltiorrhiza*, *M. guttatus* and *Se. indicum* were analyzed using this approach.

RESULTS AND DISCUSSION

Unique evolutionary history of tandem repeats in patchouli Annotation of repetitive sequences revealed a unique evolutionary history for patchouli. Although the percentage of tandem repeats in the genome was similar among four of five species investigated (Table 1), patchouli indeed had more tandem repeats and a larger genome size than the other species. The increasing percentage of tandem repeats with larger period sizes showed that the proportion of shorter tandem repeats was greater than that of longer tandem repeats in the genomes of these species (Fig. 1). Optimum matching distances with both shorter (21, 40 and 62 nt) and longer (175 and 350 nt) period sizes were discovered in the patchouli genome (Fig. 1). The three shorter period sizes (21, 40 and 62 nt) were also found in *P. cablin* reads from sequenced original reads. Similarly, in the *Se. indicum* genome, two longer period sizes (153 and 306 nt) were found, although no shorter ones were detected (Fig. 1). No specific period sizes were discovered in the genomes of *Sa. miltiorrhiza*

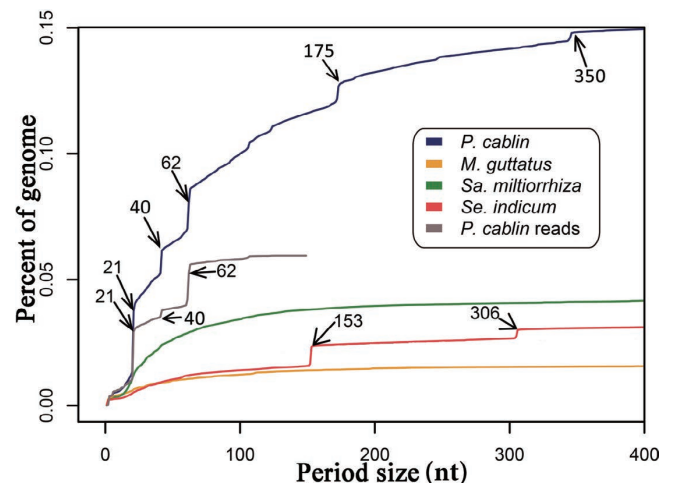


Fig. 1. Comparisons of tandem repeats from patchouli and four other species. The x-axis is the period size, and the y-axis is the percentage of the genome comprised of tandem repeats with period size equal to or less than the x-axis value. The “*P. cablin* reads” curve shows the distribution of tandem repeats identified in overlapped reads (250-bp library).

Table 1. Repeat content (TE vs. non-TE) of selected genomes

Family	Species	Accession number	Genome size (bp)	Repeat counts	Total length (bp)	Repetitive sequences (%)		
						Total	TEs	Non-TE repeats
Lamiaceae	<i>Sa. miltiorrhiza</i>	PRJNA287594	611,633,377	849,367	179,712,577	29.38	24.51	4.87
Pedaliaceae	<i>Se. indicum</i>	APMJ00000000	270,357,869	421,703	120,550,352	44.59	41.79	2.80
Lamiaceae	<i>P. cablin</i>	QKXD00000000	1,763,018,886	2,346,338	770,110,585	43.68	39.95	3.73
Phrymaceae	<i>M. guttatus</i>	APLE00000000	289,885,078	505,356	161,665,292	55.77	52.52	3.25
Lentibulariaceae	<i>U. gibba</i>	NEEC00000000	81,385,102	7,954	542,629	0.67	0.06	0.61

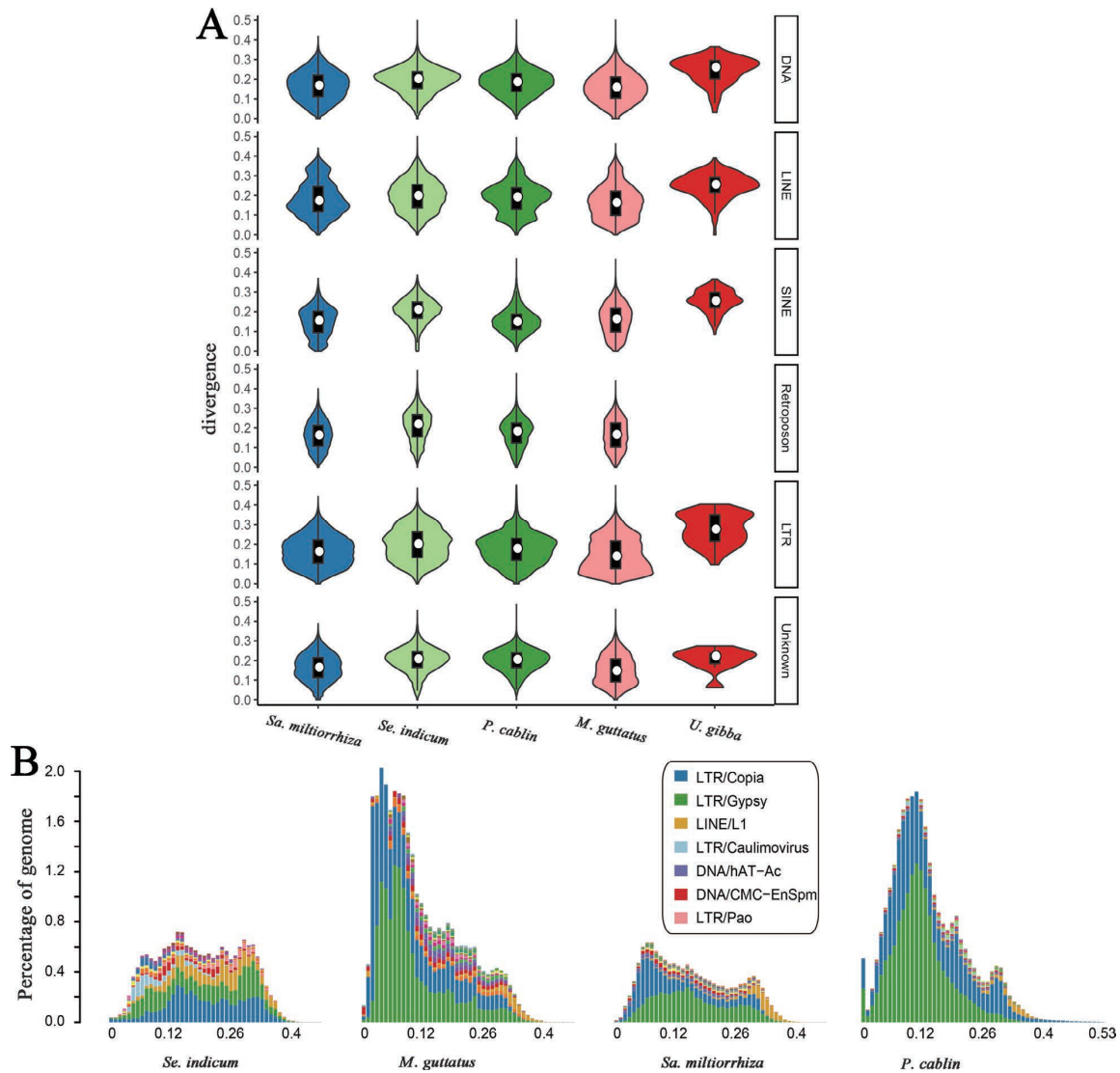


Fig. 2. Comparisons of different TEs of patchouli and other investigated species. (A) Comparisons of TEs from patchouli and other relatives. The x-axis indicates the five species investigated, and the y-axis indicates divergence of different TE families. (B) Divergence distribution of transposon sequences. Divergence is defined as the sequence divergence between a transposon sequence and the consensus sequence, which can be calculated by RepeatMasker. The x-axis includes divergence of four species, while the y-axis indicates TE percentage of genomes.

or *M. guttatus*. These optimum matching distances with period size were considered to illustrate an evolutionary history of those tandem repeats. In conclusion, these tandem repeats with smaller period size comprise a larger proportion of the patchouli genome (Fig. 1). However, their biological function remains unclear, and function should therefore be a prominent question in further investigation. Tandem repeats have been implicated in some biological functions and may also have been instrumental in biological evolution and development (Hatters and Hannan, 2013). We therefore decided to focus on this theme to interpret the characteristics of tandem repeats in the patchouli genome.

History of TE expansion Similar to other plant genomes, TEs made up a major component of repetitive sequences in the patchouli genome (Table 1). The moderate level of unknown TEs was 1.35% (Table 2), confirming that unknown repeats in patchouli and other common sequenced plants existed but did not dominate. In contrast to non-TE repeats containing assorted satellite sequences, TE repeats represented a larger proportion of the genome (more than 20%) in most plants (Table 1), because these mobile repeats have been able to spread and exist universally with a strong ability to replicate within the host genome. Moreover, the proportions of disparate TEs in these species were diverse, and a medium proportion of TE repeats appeared in the

Table 2. Repeat content by subtype in selected genomes

Sequences	<i>Sa. miltiorrhiza</i>	<i>Se. indicum</i>	<i>P. cablin</i>	<i>M. guttatus</i>	<i>U. gibba</i>
Genome size (bp)	611,633,377	270,357,869	1,763,018,886	289,885,078	81,385,102
Class I TEs					
Count (LTR) ^a	182,015	95,749	761,817	106,004	83
Length (bp, LTR)	79,773,667	49,199,479	489,383,180	81,535,693	8,262
Repeat content (% LTR)	13.04	18.20	27.76	28.13	0.01
Count (LINE)	60,632	42,446	141,579	38,998	329
Length (bp, LINE)	15,825,333	13,975,416	32,829,132	10,204,171	22,834
Repeat content (% LINE)	2.59	5.17	1.86	3.52	0.03
Count (SINE)	13,897	11,578	12,419	9,773	109
Length (bp, SINE)	1,578,334	1,326,009	1,286,444	1,396,121	7,214
Repeat content (% SINE)	0.26	0.49	0.07	0.48	0.01
Class II TEs					
Count (DNA)	265,102	196,170	768,378	255,461	119
Length (bp, DNA)	45,626,768	42,807,731	157,095,016	558,245,000	8,270
Repeat content (% DNA)	7.46	15.83	8.91	19.26	0.01
Unknown TEs					
Count (Unknown)	42,942	32,091	178,040	20,130	10
Length (bp, Unknown)	7,118,636	5,671,655	23,733,214	3,273,510	651
Repeat content (% Unknown)	1.16	2.10	1.35	1.13	0.00

^a LTR: LTR retrotransposon.

patchouli genome in comparison with other species (Table 1). Among the species investigated, most TEs in the patchouli genome had divergence values between 0.15–0.25. Compared with patchouli, *M. guttatus* possessed more TEs inserted recently, especially LTR retrotransposons, while *U. gibba* had a larger proportion of old TEs that had existed for a longer time in its genome. From the perspective of TE clusters, each of them had their own points of expanding evolution, whereas activities of most retrotransposons had relatively decentralized processes without distinct divergence peaks (Fig. 2A). In terms of the percentage divergence results, recently inserted TEs with lower divergence occupied a larger proportion of the genome of patchouli than did older ones with higher divergence (Fig. 2B). The evidence that *Copia* and *Gypsy* make up a larger proportion of the patchouli genome compared with other TEs suggests that LTR retrotransposons play important roles in the evolution of these four plants (Fig. 2B). As can be seen in Fig. 2B, *P. cablin* had an approximate span of divergence between 0.10–0.12 containing most elements, while *M. guttatus* had a span of 0.02–0.14, and these two species had more than twice as many recently inserted TEs as old ones. On the other hand, the distribution of recently inserted and old TEs in the genomes of *Se. indicum* and *Sa. miltiorrhiza* was even. At the same time, the fact that more LTR elements (more than 10%) were found than other subtypes of TEs in most of the plants investigated (Table 2) likely

illustrates the selfish roles of LTR retrotransposons without specific functions in the patchouli genome (Biémont and Vieira, 2006).

Evolutionary landscape of LTR retrotransposons

LTR retrotransposons comprised more than 60% of the repeats in patchouli, which was much higher than in other plant species investigated (Table 2), underlining the implication of a unique LTR retrotransposon evolutionary history. Indeed, the higher proportion of LTR retrotransposons with greater divergence from the consensus sequences, coupled with the higher proportion of ancient LTR retrotransposons in genome of patchouli than those of the other species, was remarkable (Fig. 2A–2B). The full-length LTR retrotransposons ranged from 3 to 9 kb in patchouli and from 5 to 10 kb in *Se. indicum* and *M. guttatus*, and from 4.5 to 7.5 kb in *Sa. miltiorrhiza* (Fig. 3A). Given that full-length LTR retrotransposons include open reading frames (ORFs) and bilateral LTRs, their ORFs accounted for the majority of the overall size of the TE (Fig. 3B). Sizes of LTRs varied from 0.25 to 1 kb in patchouli, from 0.4 to 1 kb in *Sa. miltiorrhiza*, from 0.3 to 0.7 kb in *Se. indicum* and from 0.45 to 1.35 kb in *M. guttatus* (Fig. 3B). From Fig. 4, the number of LTR retrotransposons of these four species tends to decay with time tracing back. Approximately 95% of LTR retrotransposons inserted into the genome of *M. guttatus* within the past three million years. This

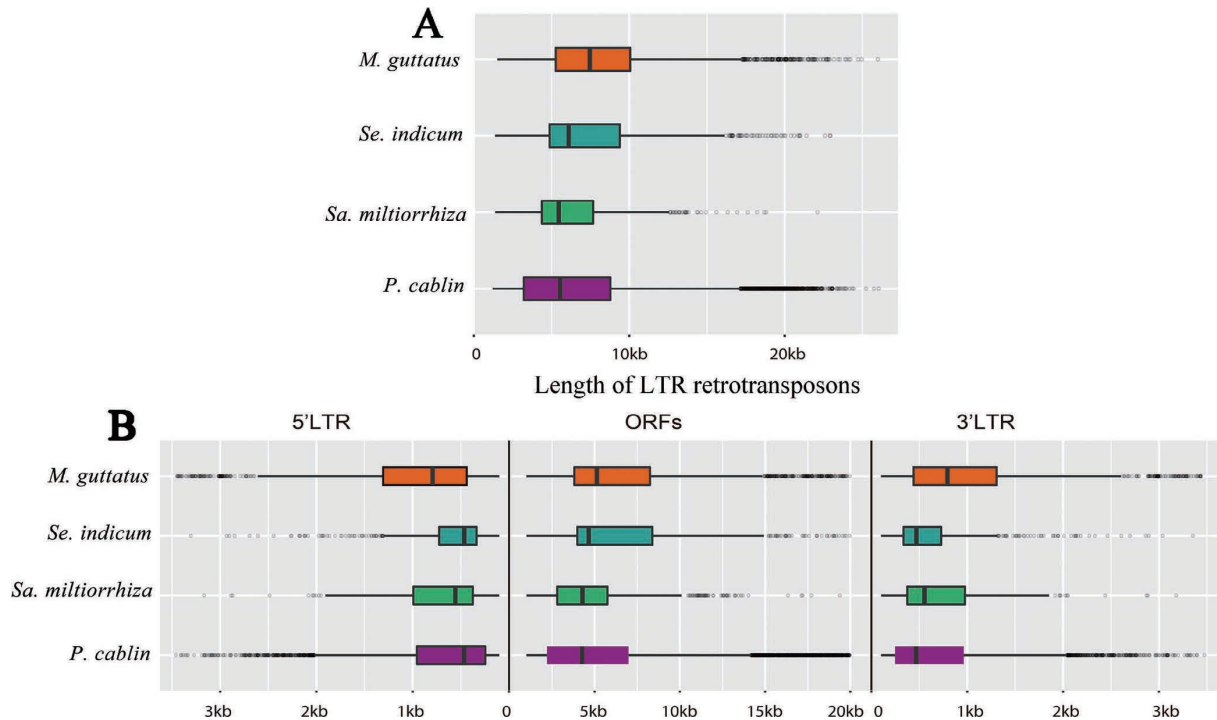


Fig. 3. Structure of LTR retrotransposons. (A) Full-length LTR retrotransposons. (B) Distribution of LTR retrotransposon components' length. The color bar and the vertical line represent length range of elements and median of lengths, respectively.

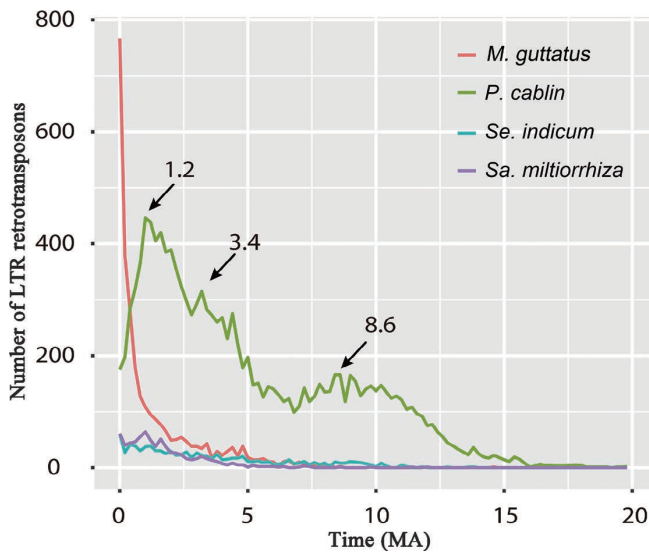


Fig. 4. Evolutionary history of LTR retrotransposons. Estimated insertion times of LTR retrotransposons (mutation rate: 8.1×10^{-9} substitutions per site per year). The x-axis is time before the present, while the y-axis is the number of LTR retrotransposons per 100 megabases of the genome. Ma, million years ago.

was mainly due to the copy/paste mechanism of LTR retrotransposons, which allows them to amplify themselves rapidly (Wicker et al., 2007). The identification of three peaks exclusively in patchouli (1.2 million years ago (Ma), 3.4 Ma and 8.6 Ma) suggested the existence of both

ancient LTR retrotransposon expansion, which happened millions of years ago, and recent expansion, although the number of recent LTR retrotransposons was twice that of the old ones (Fig. 4). Their presence in full-length form suggests that these ancient LTR retrotransposons have not yet been purged by natural selection. The reason why they have persisted may be due to the copy and paste amplification mechanism of these particular retrotransposons, which can replicate themselves continuously with the RNA-mediated mechanism. These LTR retrotransposons may be amplified rapidly when the insertion and integration of new retrotransposons occurs in the host genome, causing an increase in the number of their copies (Wicker et al., 2007). According to divergence windows, some researchers have attached importance to the repeats in certain plants, while others discovered the close relationship between repeats and evolution (Kazazian, 2004). Nevertheless, little work had been performed on this herb called patchouli. To rectify this oversight, this study focused on both non-TE repeats and TE repeats classified as class I and class II, such as DNA TEs, LTR retrotransposons, short interspersed nuclear elements and long interspersed nuclear elements, to explore the evolution of patchouli.

This study connected patchouli to evolutionary studies at the genomic level using bioinformatic methods. Cumulatively, our research should be of great value to the further detection and understanding of the evolutionary history of repeats. However, it should be noted

that very little optimization work has been performed on functions in our study. For instance, despite the compelling evidence that patchouli indeed has more tandem repeats than other species except *Sa. miltiorrhiza*, biological features of these repeats have not yet been discovered. Hence, an additional functional component study integrated with genome-analyzing technology will be valuable for further investigations. It is feasible to use our methods to explore the similarities and homology, as well as to make discoveries of potential value, of this plant and to compare the divergence between particular sequences of several plant relatives. This study thus offers a new strategy to utilize this herb.

This work was supported by Grant No. 2017JQ0015 from the Outstanding Youth Science Foundation of Sichuan Province.

REFERENCES

- An, X., Wu, W. R., Lai, H. L., Yang, L., Tan, X. N., and Huang, Z. B. (2019) Research progress in molecular pharmacognosy of *Pogostemon cablin*. *China J. Chin. Mater. Med.* **44**, 4781–4785 (in Chinese).
- Bennetzen, J. L. (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr. Opin. Genet. Dev.* **15**, 621–627.
- Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580.
- Beulé, T., Agbessi, M. D., Dussert, S., Jaligot, E., and Guyot, R. (2015) Genome-wide analysis of LTR-retrotransposons in oil palm. *BMC Genomics* **16**, 795.
- Biémont, C., and Vieira, C. (2006) Genetics: junk DNA as an evolutionary force. *Nature* **443**, 521–524.
- Biscotti, M. A., Olmo, E., and Heslop-Harrison, J. S. (2015) Repetitive DNA in eukaryotic genomes. *Chromosome Res.* **23**, 415–420.
- Britten, R. J., and Kohne, D. E. (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**, 529–540.
- Bucher, E., Reinders, J., and Mirouze, M. (2012) Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr. Opin. Plant Biol.* **15**, 503–510.
- Chinese Pharmacopoeia Commission (2010) Pharmacopoeia of the People's Republic of China 2010. China Medical Science Press, Beijing, China.
- Du, J., Tian, Z., Hans, C. S., Laten, H. M., Cannon, S. B., Jackson, S. A., Shoemaker, R. C., and Ma, J. (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J.* **63**, 584–598.
- Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405.
- Feschotte, C., Jiang, N., and Wessler, S. R. (2002) Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* **3**, 329–341.
- Finnegan, D. J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**, 103–107.
- Hatters, D. M., and Hannan, A. J. (2013) Tandem Repeats in Genes, Proteins, and Disease: Methods and Protocols. Humana Press, Totowa, USA.
- He, Y., Peng, F., Deng, C., Xiong, L., Huang, Z.-y., Zhang, R.-q., Liu, M.-j., and Peng, C. (2018) Building an octaploid genome and transcriptome of the medicinal plant *Pogostemon cablin* from Lamiales. *Sci. Data* **5**, 180274.
- He, Y., Xiao, H., Deng, C., Xiong, L., Nie, H., and Peng, C. (2016) Survey of the genome of *Pogostemon cablin* provides insights into its evolutionary history and sesquiterpenoid biosynthesis. *Sci. Rep.* **6**, 26405.
- Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., Schmutz, J., Willis, J. H., and Rokhsar, D. S. (2013) Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc. Natl. Acad. Sci. USA* **110**, 19478–19482.
- Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481.
- Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C., Carretero-Paulet, L., Chang, T.-H., Lan, T., Welch, A. J., Juárez, M. J. A., Simpson, J., et al. (2013) Architecture and evolution of a minute plant genome. *Nature* **498**, 94–98.
- Kazazian, H. H., Jr. (2004) Mobile elements: Drivers of genome evolution. *Science* **303**, 1626–1632.
- Kidwell, M. G., and Lisch, D. R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**, 1–24.
- Kumar, A., and Bennetzen, J. L. (1999) Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532.
- Lan, T., Ranner, T., Ibarra-Laclette, E., Farr, K. M., Chang, T.-H., Cervantes-Pérez, S. A., Zheng, C., Sankoff, D., Tang, H., Purbojati, R. W., et al. (2017) Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc. Natl. Acad. Sci. USA* **114**, 4435–4441.
- McCarthy, E. M., and McDonald, J. F. (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**, 362–367.
- Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., Zhang, Y., Zhang, X., Wang, Y., Hua, W., et al. (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* **15**, R39.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.
- Wu, Y. G., Guo, Q. S., and Zheng, H. Q. (2007) Textual research on history of introduction and herbal medicine of *Pogostemon cablin*. *China J. Chin. Mater. Med.* **32**, 2114–2117 (in Chinese).
- Yin, H., Du, J., Wu, J., Wei, S., Xu, Y., Tao, S., Wu, J., and Zhang, S. (2015) Genome-wide annotation and comparative analysis of long terminal repeat retrotransposons between pear species of *P. bretschneideri* and *P. Communis*. *Sci. Rep.* **5**, 17644.
- Zhang, G., Tian, Y., Zhang, J., Shu, L., Yang, S., Wang, W., Sheng, J., Dong, Y., and Chen, W. (2015) Hybrid *de novo* genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). *GigaScience* **4**, 62.