Complete chloroplast genome sequence of *Amomum villosum* and comparative analysis with other Zingiberaceae plants

Li Yang a, Chong Feng a, Miao-miao Cai a, Jie-hu Chen b, Ping Ding a,⇑

a School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510006, China
b Science Corporation of Gene, Guangzhou 510000, China

Abstract

Objective: *Amomum villosum* (AV) is an herb whose dried fruit has been extensively used in modern medicine to treat digestive system diseases such as dysentery, vomiting and abdominal pain. This paper aims to supplement chloroplast (cp) genomic resources and to be used in phylogenetic studies and identification of AV related plants.

Methods: High-throughput sequencing technology was used to determine the complete sequence of the AV cp genome, and the sequence was then compared with three related species.

Results: The genome size of AV we obtained was 163,968 bp with an obvious tetrad structure. The AV cp genome was observed to contain 125 unique genes and 81 simple sequence repeat (SSRs) had been determined and the majority of which were adenine–thymine (AT)-rich. Comparative analysis of genome sequence of four ginger plants showed that the *atpF*, *clpP* and *rpl32* genes are potential markers for identifying *Amomum* species. Phylogenetic analysis suggested that AV was closely related to *A. kravanh* and *A. compactum*.

Conclusion: These results have brought useful genetic resources for further identification researches, DNA barcoding, resolving taxonomy and understanding the evolutionary mode of Zingiberaceae cp genome.

© 2020 Tianjin Press of Chinese Herbal Medicines. Published by ELSEVIER B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The chloroplast (cp) is formed by endosymbiotic interactions between photosynthetic bacteria and non-photosynthetic hosts (Xiang et al., 2016). The cp plays important roles throughout the plant life cycle, including photosynthesis, metabolism, and starch, pigments, fatty acids, and amino acids biosynthesis (Daniell, Lin, Yu, & Chang, 2016; Liu et al., 2018; Park et al., 2017c). Due to containing valuable information with highly conservative nature, the cp genome has been widely used in molecular markers, barcode identification, phylogenetic analysis and other fields (Wu et al., 2017). The cp genomes exhibit a circular quadripartite structure ranging from 120 to 160 kb in length, and these genomes typically consist of four parts: one long single-copy (LSC) region, one short single-copy (SSC) region, and two copies of a large inverted repeat (IR) region.

The cp genome exhibits a typical tetragonal structure with a length between 120 and 160 kb. These genomes are usually composed of four parts: a long single copy area (LSC), a short single copy area (SSC) and two reverse repeat regions (IR) (Choi and Park, 2015; Yu et al., 2017). The number of genome sequences was significantly increased because of the next-generation sequencing technologies development (Wu et al., 2017). As of February 2019, 2093 complete cp genomes of land plants had already been added to the GenBank, which enable us to gain insights into plant biological diversity, DNA barcoding, evolution, and population genetic analysis (Daniell et al., 2016; Benson et al., 2018; Baczkiewicz et al., 2017; Song et al., 2017).

*Amomum villosum* Lour. (AV, family Zingiberaceae) is a valuable herbaceous plant distributed in Southeast Asia (e.g., Burma, Laos, and especially in Southern China) (Wang et al., 2018; Li et al., 2010). Its medicinal parts are the ripe fruits or seed groups, which mainly contains volatile terpenes and has antibacterial, anti-ulcer, and anti-diarrhea activities (Huang et al., 2014; Xue et al., 2015; He et al., 2015; Chen et al., 2018). Due to its aroma and flavor, AV can also be used as culinary spices to prepare beverages, tea and some foods (Wang et al., 2018). The wide use of AV has increased demand for the fruits, which are mainly produced from cultivation. AV relies only on artificial pollination during the planting stage because of its flowers’ special structure, where the stigma of the pistil is higher than the stamen anther, making natural pollination difficult (He et al., 2014). Its yield is extremely low (36–60 kg/acre),...
with high market price of about 3000–5000 RMB/kg. Therefore, there are a number of adulterants and counterfeit AV in the market. Some related fruits or seeds are usually exchanged with AV, such as Amomum compactum Soland. ex Maton, Amomum kravanh Pierre ex Gagnep, Alpinia oxyphylla Miquel, and others, causing consumer health safety hazards and concerns (Wang et al., 2000). These factors have seriously affected the quality of AV. There is a high degree of complexity and diversity in morphology and internal structure, relying on traditional methods to identify them are more difficult. Although there have been some researches related on molecular identification and could provide some information about the taxonomy of AV and its related species (Wang et al., 2000; Wu et al., 2018; Pan et al., 2001; Zhang et al., 2018), few studies were performed on the genetic diversity of cp genome in AV. Therefore, there is a need to apply new methods to identify the AV and adulterants and abundant AV gene resources.

In this paper, the cp DNA structure of AV was determined and analyzed, including its essential organization, codon usage, and comparison of the entire genome. The phylogenetic tree was then constructed using the protein-coding genes of 11 plants in three genera of Zingiberaceae. Our results provide a complete AV cp genome, which is beneficial to phylogenetic research, breeding and identification of the plants related to AV.

2. Materials and methods

2.1. Leaf DNA extraction and sequencing

The fresh AV leaves were gathered from Yangchun City, which is located in Guangdong Province, China, well-known for producing genuine and high-quality AV. The samples were identified by Prof. Ping Ding (Guangzhou University of Chinese Medicine, Guangdong, China). Voucher specimens were deposited in the herbarium of Guangzhou University of Traditional Chinese Medicine, China. Total DNA was extracted from liquid nitrogen ground leaf powders using a Plant Genomic DNA Kit (Zhanchen Biotech Co., Guangzhou, China) (Wang et al., 2016). The quality and integrity of DNA samples were tested by NanoDrop 2000 spectrometer (Thermo Scientific, Waltham, MA, USA) and agarose gel electrophoresis (Saina et al., 2018). High-quality cpDNA was used to prepare 500 bp (insert size) pair-end DNA sequencing according to the manuscript library (Yang et al., 2018), and were sequenced by employing the Illumina Hiseq 4000 platform (Illumina Inc., San Diego, CA, USA).

2.2. Genome assembly

Briefly, filtering the sequencing reads based on quality value, and the bases which quality < 20 and error rate > 0.01 of 3' downstream and 5' upstream were clipped (Yang et al., 2018). First, pair-end sequencing reads were de novo assembled using SOAPdenovo 2 (http://soap.genomics.org.cn/soapdenovo.html) with multikmer (35–75). Second, all reported cp genome sequences of dicot were referenced, and all contigs were aligned using BLAST+ (National Center for Biotechnology Information, Bethesda MD, USA) by using the blastn method (Wang et al., 2016). Third, we assembled the contigs to the genome with overlap and read the pair-end relationship. Additionally, the polymerease chain reaction (PCR) was carried on to examine the assembly between LSC/SSC and IRs areas (Table S1).

2.3. Gene annotation and codon usage

The CPGAVAS (Chinese Academy of Medical Sciences, Beijing, China) and DOGMA (University of Texas at Austin, Austin, TX, USA) were used for preliminary gene annotation (Wu et al., 2017). The rRNA genes were confirmed using blastn with a nt database (Yang et al., 2018; Cheng et al., 2013; Iwasaki et al., 2013). The tRNAscan-SE v.2.0 (University of California Santa Cruz, CA, USA) software was used to verify the tRNA genes (Schattner et al., 2005) and the OGCDRAW (the Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany) program was used to generate a circular genome map. The final cp genome of AV was saved in GenBank with the accession number (pending upload). The GC contents and relative synonymous codon usage values (RSCU) of AV genome were analyzed using MEGA7 software to character codon usage (Kumar et al., 2016). The RSCU was the ratio between the actual observed value and the theoretical observed value of the codon (Gu et al., 2018).

2.4. Identification of long repetitive sequences and simple sequence repeats analysis

The size and position of long repeats in the cp genome of three species, including forward repeats, inverted repeats, palindrom repeat, and complement repeats were determined using REDuter software with following parameters: the minimum repeat length = 20 bp, sequence similarity > 90% and the Hamming distance = 3. The positions and types of simple sequence repeats (SSRs) within the AV genome were identified by the MISA software (Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany). The microsatellites are some tandem repeats with one to six nucleotides distributed throughout the genome. The SSR thresholds were 10, 6, 5, 5 and 5 for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentonucleotide and hexonucleotide, respectively (Wang et al., 2016).

2.5. Analysis of synonymous and non-synonymous substitution rates

Using DnaSP 5 method (The University of Barcelona, Barcelona, Spain) to estimate the ratio of synonymous and (Ka) non-synonymous (Ks) substitution for protein-coding gene between the AV cp genomes and other two plants were estimated (Librado & Rozas, 2009). In order to evaluate the mutation rate of Ka/Ks, a Python script was carried out to extract the shared single protein coding exons and align them with MEGA7 (Wu et al., 2017).

2.6. Cgview comparison tool (CCT) map

The CCT method (University of Alberta, Alberta, Canada) was used to compare AV cp genome with other available plants of Zingiberaceae (Kyalo et al., 2018), and the result was shown as a circle. We signed genes by orthologous groups clusters, and the BLAST software was used to compare AV with other genomes. The distributions of AT were analyzed based on AT skew as follows: AT skew = [A - T]/[A + T].

2.7. Phylogenetic analysis

A total of 77 common protein-coding genes from 12 species cp genomes were used to determine the AV phylogenetic location, including an outlier plant (T. latifolia). jModeltest 0.1.1 software (The University of Vigo, Vigo, Spain) was applied based on the Akaike information criterion (AIC) to analyze the model of GTR + G + I for the nucleotide sequence (Wang et al., 2016), followed by building the phylogenetic trees using RAxML 8.1.5 software with a rapid bootstrap analysis (1000 replicates) (Stamatakis, 2014). Phylobayes 4.1b was performed to Bayesian inference (BI) analysis with two chain max diff < 0.01 (Wang et al., 2016).
3. Results and discussion

3.1. Features of AV cp genome

We determined the complete cp genome of AV, a typical quadri-partite structure, to be 163,968 bp in size. The large (LSC; 88,798 bp) and short (SSC; 15,352 bp) single copy regions were split by two inverted repeats (IRs; 29,909 bp) (Table 1 and Fig. 1). The overall GC content of the AV genome was 36.58%, with the IR regions possessing higher GC content (41.08%) than the LSC (33.72%) and SSC regions (29.99%) due to the reduction of AT nucleotides in the four duplicate rRNA genes. Within the protein-coding regions (CDS), the adenine-thymine (AT) content of the third-codon positions (71.2%) was higher than that of the first

Table 1

| Region   | A(U)% | T/%  | G/%  | C/%  | Length /bp |
|----------|-------|------|------|------|------------|
| LSC      | 32.48 | 33.80| 16.50| 17.22| 88,798     |
| SSC      | 35.79 | 34.23| 14.27| 15.72| 15,352     |
| IRA      | 30.20 | 28.72| 21.26| 19.82| 29,909     |
| IRB      | 28.72 | 30.20| 19.82| 21.26| 29,909     |
| Total    | 31.69 | 32.26| 18.29| 18.29| 163,968    |
| 1st position | 31.50 | 23.97| 26.35| 18.19| 27,757     |
| 2nd position | 30.20 | 32.37| 17.38| 20.05| 27,757     |
| 3rd position | 32.83 | 38.37| 15.50| 13.30| 27,757     |

LSC: long single-copy; SSC: short single-copy; IR: inverted repeat; CDS: protein-coding regions; A: adenine; T: thymine; G: guanine; C: cytosine.

Fig. 1. Complete chloroplast (cp) genome map of *A. villosum* (AV). Gene drawn inside circle is transcribed clockwise, while outside is just opposite. Color coding of genes is based on functional groups they belong to. Dark gray color of inner circle indicates GC content.
In total, the cp gene of AV has 125 unique functional genes, including 87 protein-coding genes (PCGs), 30 transfer RNA (tRNAs) genes, and eight ribosomal RNA (rRNAs) genes (Table 2). Of these, 18 genes were present as duplicates: six tRNAs, four rRNAs, and eight protein-coding genes (\textit{rps12}, \textit{rps19}, \textit{rpl2}, \textit{rpl23}, \textit{ndhA}, \textit{ndhB}, and \textit{ycf2}). All eight rRNA were situated in IR regions (Liu et al., 2018; Liu, Yang, Zhao, Li, & Xiang, 2017; Zhou et al., 2018), which is consistent with numerous research results. The \textit{rps19} gene was found to be located in the IR and LSC boundary region, and \textit{ycf1} gene was mapped in the junction of IR and SSC.

Introns play an inseparable role in regulating gene expression, and can increase the expression of foreign genes at a specific time and location, they therefore can be used as an important tool to improve the efficiency of transformation (Yi et al., 2012). Among the 125 functional genes, eight genes contained introns and most of them contained only one, whereas \textit{ycf3} and \textit{clpP} harbored two, similar to other species (Chen et al., 2015; Curci et al., 2015).

Among the eight intron genes, five protein-coding genes were mapped at the LSC, two at the IR, and only one gene at the SSC area (Table 3). In particular, the \textit{rps12} gene was a trans-spliced gene, in which 5' exon was located at the LSC area, while the 3' exon and intron were replicated in the IR area. The \textit{ndhA} gene contained the longest intron region (1049 bp).

### Table 2
**Gene contents in \textit{A. villosum} cp genome.**

| Gene category | Gene groups | Gene names |
|---------------|-------------|------------|
| Self-replication | Transfer RNAs | 30 tRNA genes |
| Ribosomal RNAs | \textit{rrn16} (\textit{c}2), \textit{rrn23} (\textit{c}2), \textit{rrn4.5} (\textit{c}2), \textit{rrn5} (\textit{c}2) |
| Ribosomal proteins (SSU) | \textit{rps2}, \textit{rps3}, \textit{rps4}, \textit{rps7} (\textit{c}2), \textit{rpS8}, \textit{rpS11}, \textit{rpS12} (\textit{c}2), \textit{rpS14}, \textit{rpS15}, \textit{rpS16}, \textit{rpS18}, \textit{rpS19} (\textit{c}2) |
| Ribosomal proteins (LSU) | \textit{rpL2} (\textit{c}2), \textit{rpL14}, \textit{rpL16}, \textit{rpL20}, \textit{rpL22}, \textit{rpL23} (\textit{c}2), \textit{rpL32}, \textit{rpL33}, \textit{rpL36} |
| Photosynthesis | NADH-dehydrogenase | \textit{ndhA} (\textit{c}2), \textit{ndhB} (\textit{c}2), \textit{ndhC}, \textit{ndhD}, \textit{ndhE}, \textit{ndhF}, \textit{ndhG}, \textit{ndhH}, \textit{ndhI}, \textit{ndhJ}, \textit{ndhK} |
| Photosystem I | \textit{psaA}, \textit{psaB}, \textit{psaC}, \textit{psaL}, \textit{psaL} |
| Photosystem II | \textit{psbA}, \textit{psbB}, \textit{psbC}, \textit{psbD}, \textit{psbE}, \textit{psbF}, \textit{psbH}, \textit{psbl}, \textit{psbJ}, \textit{psbK}, \textit{psbM}, \textit{psbN}, \textit{psbT}, \textit{psbZ} |
| Cytochrome b/f complex | \textit{petA}, \textit{petB}, \textit{petD}, \textit{petG}, \textit{petI}, \textit{petN} |
| ATP synthase | \textit{atpA}, \textit{atpB}, \textit{atpC}, \textit{atpE}, \textit{atpF}, \textit{atpH}, \textit{atpL} |
| Proteins | Rubisco Large subunit | \textit{rbcL} |
| Other | Translational initiation factor | \textit{infA} |
| Maturation | \textit{matK} |
| Protease | \textit{clpP} |
| Envelope membrane protein | \textit{cemA} |
| Subunit of Acetyl-CoA-carboxylase | \textit{accD} |
| \textit{c-Type cytochrome synthesis gene} | \textit{ccoA} |
| Hypothetical chloroplast reading frames (\textit{ycf}) | \textit{ycf1}, \textit{ycf2} (\textit{c}2), \textit{ycf3}, \textit{ycf4} |

Note: (\textit{c}2) Genes with two copies.

### Table 3
**Genes with introns in AV cp genome, and lengths of exons and introns.**

| Genes | Locations | Exon I /bp | Intron I /bp | Exon II /bp | Intron II /bp | Exon III /bp |
|-------|-----------|------------|--------------|-------------|---------------|--------------|
| \textit{atpF} | LSC | 390 | 796 | 168 |
| \textit{rpoc1} | LSC | 1638 | 723 | 423 |
| \textit{ycf3} | LSC | 153 | 777 | 228 | 715 | 152 |
| \textit{clpP} | LSC | 352 | 630 | 1300 | 841 | 69 |
| \textit{rpl2} | IR | 435 | 659 | 384 |
| \textit{ndhB} | IR | 756 | 700 | 777 |
| \textit{ndhA} | SSC | 540 | 1049 | 555 |
| \textit{rps12} | LSC/IR | 114 | – | 114 |
Delannoy et al., 2018), and is the main reason for the relative conservation of cp genes. Furthermore, the usage of the start codons AUG (encoding Met) and TGG (encoding Trp) exhibited no bias (RSCU = 1).

3.3. Repeat and simple sequence repeats analysis

Repeated regions are of great significance in the evolutionary process, and can affect changes in genome structure like substitution and duplication, and they mostly occur in the sequences of intron and intergenic spacer (IGS) (Park, Kim, Yeo et al., 2017). In this study, we used REDuter to compare the forward (F), palindrome (P), reverse (R), and tandem (T) types (>30 bp) of the three plants of Amomum, and found that the AV genome had the largest number of repeats (47F, 39P, 12 R, and 12 T types), followed by A. compactum (41F, 34P, 20 R and 14 T types), while A. kravanh had the least repeats. Among them, the F and P types accounted for the largest proportion with lengths mainly varied from 20 to 39 bp (Fig. 3A and B). The generation of F types is often related to the activity of transposons, which can lead to variations in genome structure, and is usually used as a marker for population relationship studies (Gu et al., 2018). There were 69.09% repeats located in the intergenic area, 17.27% in coding area, and 13.64% of the sites such as atpF and rps12 were located in the intron area (Fig. 3C).

Simple sequence repeat (SSR), or microsatellite was extensively used in bioengineering, breeding and phylogenetic research as an effective technical method (Zhang et al., 2016). There were 81 SSRs have been identified, and mononucleotides had the largest number, accounting for 88.89% of the total SSRs, followed by dinuc-
cleotide SSRs (Fig. 4A). However, the tetra-, penta-, or hexa-nucleotide SSRs couldn’t be found according to our method. The mononucleotide repeats were mainly composed of A and T, which may lead to A/T richness in the angiosperm cp genomes (Liu et al., 2018). The results are similar to the statement that, is, most SSRs are consisted of short poly A (polyadenine) or poly T (polythymine) sequences. Approximately 13.58%, 9.88% and 76.54% SSRs were presented in protein-coding, introns, and intergenic areas of the AV, respectively (Fig. 4B). These findings are similar, as the distribution of SSR is unbalanced in the genome (Park et al., 2017).

3.4. Analysis of synonymous and non-synonymous substitution rates

The non-synonymous (Ka) and synonymous (Ks) substitution ratio (denoted as Ka/Ks) is an important tool, mainly used to estimate the evolutionary pressures in specific groups of genes. Ratios > 1 indicate positive selection, values < 1 (especially if < 0.5) indicate negative selection, and values close to 1 indicate neutral selection (Wu et al., 2017). Here, we analyzed the Ka/Ks ratio of the 79 unique protein-coding genes in AV, A. compactum and A. kravanh genome (Table S3). Of these, most of the proteins possessed Ka/Ks ratios < 0.5, which suggests that most of protein-coding genes faced great pressure for purification and selection. In atpF and ycf1, the Ka/Ks values were > 1, which indicated a positive selection. In the genes involved in photosynthesis, for example, the atpF gene can encode a subunit of the H+-ATP synthase, thereby affecting the electron transport and photosynthetic phosphorylation in photosynthesis. These findings suggested that different levels of selective pressure in species may affect the function of the cp gene (Wu et al., 2017).

3.5. CG view comparison tool (CCT) map

Four available cp genomes of Zingiberaceae species (A. compactum, A. kravanh, A. oxyphilla, and Curcuma flaviflora S Q. Tong) were selected for comparison with AV because the former is morphologically similar to AV (Fig. S2). The sequence identity between AV and other species’ cp genomes was analyzed using CGView (University of Alberta, Alberta, Canada) with the annotated AV sequence as the reference. The closeness between plants generally reflects the similarity of gene sequences. The results revealed that the sequence similarity of A. compactum genome was the highest (>90%), followed by A. kravanh, similar to the cluster analysis results. The most similar regions were located in the IR area, and the SSC and LSC areas were quite different among these genomes involved in this study.

Differences in coding regions were smaller than non-coding regions, and the region with the greatest divergent lies in the intergenic area (Wu et al., 2017). The most divergent genes between AV and A. compactum were atpF, clpP, and rpl32, for which the blast identity values were 98.03, 98.87, 98.85, respectively (Fig. 5). The greatest difference in genes between AV and A. kravanh were atpF and rpl32, which were also considered to have high variability in other species (Yin et al., 2018). These three genes can be better used in the identification of Amomum family, even other species. Hence, the atpF, clpP, and rpl32 genes may be considered for development as molecular markers and barcoding to differentiate Amomum species. Among them, the atpF gene was also strongly positively selected. Above results bring a new insight for the development of molecular markers for Amomum family and even other species.

3.6. Phylogenetic analysis

The cp sequences are often used in phylogenetics, evolution, and molecular systems studies (Liu et al., 2018). To determine the phylogenetic relationship of AV in Zingiberales, we used Maximum (ML) and Bayesian (BI) nucleic acid to analyze 77 protein-coding genes commonly found in 12 plants including AV. (Fig. 6). The results of ML and BI exhibited similar phylogenetic topologies. All nodes with a Bootstrap value of 100% were found using ML, and nine of them had observed bootstrap values ≥ 95% based on BI. Similarly, ML and BI protein analyses revealed that 8 of 9 nodes with bootstrap values of ≥ 99% (Fig. 7). Both nucleic acid and protein analyses showed that Amomum and Alpinia plants were sister groups. The four plants of AV, A. compactum, A. kravanh, and A. oxyphilla were grouped with 100% bootstrap values, and AV clustered more closely with A. compactum and A. kravanh than with A. oxyphilla.

The results of our cluster analysis are basically consistent with the phenotype-based clustering results reported in the literature (Zhang, 1994; Benedict et al., 2015). A. compactum and A. kravanh are plants from the same genus as AV, and are similar in the exterior shape and interior structure of their fruits. They are the most common counterfeits in many markets. A. oxyphilla is part of a related genus plants, and its seed groups are similar to AV’s (Fig. 8) and it may be occasionally used as AV. The chemical constituents of the above four plants have considerable differences (Ding et al., 2004), and misuse may pose a threat to human health.

Fig. 4. Simple sequence repeats (SSRs) analysis in AV cp genome. (A) Frequency distribution of different classes of polymer in cp genome of AV and (B) SSRs frequency identified in intergenic spacer (IGS), protein-coding (CDS) and intron regions.
Fig. 5. Genome comparison of four Zingiberaceae cp genomes to AV. Species involved are *A. oxyphylla*, *A. compactum*, *C. flaviflora*, and *A. kravanh* from outside. Four outermost rings represent protein-coding locations, while inner two rings indicated adenine–thymine (AT) skew. "AT skew +" indicates A > T, "AT skew −" indicates A < T.

Fig. 6. Genetic relationship of AV based on maximum likelihood (ML) and bayesian inference (BI) nucleic acid analyses of 77 genes. First number represents BI bootstrap value of each branch, and the last one corresponds to ML. Phylogenetic tree was drawn using *Typha latifolia* as an outgroup. Position of AV is shown in boldface.
The results will facilitate the use of molecular markers to identify species of AV and other genera.

4. Conclusion

The AV entire cp genome was analyzed in this study, and the genome obtained had a quadruple structure. There were 81 SSRs in the AV genome, which will be used for further species identification. The ratio of Ka/Ks revealed that a large proportion of genes were in a state of strong purification selection. The fruits or seeds of A. kravanh, A. compactum and A. oxyphylla are usually substituted for AV due to highly similar morphological traits, and the phylogenetic tree fully supported AV as being closely related with A. kravanh and A. compactum, with a 100% bootstrap value.

Comparison of cp genomes of the three plants indicated that the differences between them are very slight, while the atpF, clpP, and rpl32 genes are the most highly divergent regions, which will be developed as molecular markers that could discriminate the above related plants. The cp genome information of AV is an essential genetic resource that may facilitate the molecular identification of AV, and will lay a way for the breeding of a good cultivar of AV.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key R&D Program of China (2017YFC1701104) and Guangdong Province Applied Science and Technology R&D Special Fund Project (2015B020234002).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chmed.2020.05.008.

References

Baczkiewicz, A., Szczechinska, M., Sawicki, J., Stebel, A., & Buczkowska, K. (2017). DNA barcoding, ecology and geography of the cryptic species of Aneura pinguis and their relationships with Aneura musista and Aneura mirabilis (Metzgeriales, Marchantiophyta). *PLoS ONE, 12*(12), e0188837.
Benedict, J. C., Smith, S. Y., Collinson, M. E., Specht, C. D., Marone, F., Xiao, X. H., & Parkinson, D. Y. (2015). Seed morphology and anatomy and its utility in recognizing subfamilies and tribes of Zingiberaceae. *American Journal of Botany, 102*(11), 1814–1841.
Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Rapp, B. A., & Wheeler, D. L. (2018). *Genbank. Nucleic Acids Research, 46*(D1), D41–D47.
Cheng, J., Zeng, X., Ren, G., & Liu, Z. (2013). *CcAP: A new comprehensive platform for the comparative analysis of chloroplast genomes, BMC Bioinformatics, 14*, 95.
Chen, J. H., Hao, Z. D., Xu, H. B., Yang, L. M., Liu, G. X., & Sheng, Y. (2015). The complete chloroplast genome sequence of the relict woody plant Metasequoia glyptostroboids Huet Cheng. *Frontiers in Plant Science, 6*, 447.
Chen, Z., Ni, W. Y., Yang, C. X., Zhang, T., Lu, S. H., Zhao, R. H., & Mao, X. J. (2018). Therapeutic effect of Amomum volsum on inflammatory bowel disease in rats. Frontiers in Pharmacology, 9, 639.

Choi, K. S. & Park, S. (2015). The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae): Genomic features and relationship with Asteraceae. *Gene*, 57(2), 214–221.

Curci, P. L., De P. D., Vendramin, G. C., & Sonnante, G. (2015). Complete chloroplast genome of the multifructocarp globe Artichoke and comparison with other Asteraceae. *PLoS One*, 10(3) e0120589.

Danniel, H., Lin, C., Yu, M., & Chang, W. (2016). Chloroplast genomes: Diversity, evolution and applications in genetic engineering. *Genome Biology and Evolution*, 8 (7), 2077–2086.

Ding, P., Fang, Q., & Xu, H. H. (2004). GC fingerprint of components of *Amomum cornutum* (Gingkoaceae). *Molecular Biology and Evolution*, 11 (1), 99–102.

He, G. Z., Gao, W., Su, J. K., Li, J. K., & Tang, L. Y. (2014). Morphological characteristics of the medicinal plant *Amomum volsum* flower organ. *Chinese Bulletin of Botany*, 49 (3), 313–321.

He, X. Y., Wang, H., Yang, J. F., Deng, K., & Wang, T. (2018). RNA sequencing on *Dingoe*, *P. rhyneina* and *D. macrosorum* (Asteraceae). *Plant Science*, 81, 174–182.

Danniel, H., Lin, C., Yu, M., & Chang, W. (2016). Chloroplast genomes: Diversity, evolution and applications in genetic engineering. *Genome Biology and Evolution*, 8 (7), 2077–2086.

Ding, P., Fang, Q., & Xu, H. H. (2004). GC fingerprint of components of *Amomum volsum* and its counterfeit species. *West China Journal of Pharmaceutical Sciences*, 19(5), 330–332.

Gu, C. H., Dong, B., Xu, L., Tembrock, L. R., Zheng, S. Y., & Wu, Z. Q. (2018). The complete chloroplast genome of *Heinolia myrtifolia* and comparative analysis within Myrtales. *Molecules*, 23(4), 846.

He, G. Z., Gao, W., Su, J. L., & Tang, L. Y. (2014). Morphological characteristics of the medicinal plant *Amomum volsum* flower organ. *Chinese Bulletin of Botany*, 49 (3), 313–321.

He, X. Y., Wang, H., Yang, J. F., Deng, K., & Wang, T. (2018). RNA sequencing on *Amomum volsum* Lour. induced by MeJa identifies the genes of WKY and terpenoids synthesis in terpene biosynthesis. *Genome*, 61(2), 91–99.

Huang, Q., Duan, Z., Yang, J., Ma, X., Zhan, R., & Xu, H. (2014). SNP typing for germlapse identification of *Amomum volsum* Lour. based on DNA barcoding markers. *PLoS One*, 9(12) e114940.

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., & Sado, T. (2013). Mitofish and mitoannotator: A mitochondrial genome database of fish and mammalian mitogenomes and a pipeline for automatic genome annotation. *PLoS One*, 8(7), e63990.

Librado, P., & Rozas, J. (2009). DanSP v5: A software for comprehensive analysis of SNP data. *Molecular Biology and Evolution*, 26(4), 895–897.

Li, G., Chen, A. J., Chen, X. Y., Li, X. L., & Gao, W. W. (2010). First report of *Amomum volsum* (cardamom) leaf lesion caused by *Pyricularia costina*. *China Plant Disease Reports*, 22(7), 681–685.

Liu, X., Zhou, B. Y., Yang, H. Y., Li, Y., Yang, Q., Lu, Y. Z., & Gao, Y. (2018). Sequencing and characterization of *Chrysanthemum carinatum* Schousb and *Kalimeris indica* L. *European Journal of Horticultural Science*, 92(2), 1–10.

Liu, X., Zhou, B. Y., Yang, H. Y., Li, Y., Yang, Q., Lu, Y. Z., & Gao, Y. (2018). Sequencing and characterization of *Chrysanthemum carinatum* Schousb and *Kalimeris indica* L. *European Journal of Horticultural Science*, 92(2), 1–10.

Kumar, S., Stecher, G., & Tamura, K. (2016). Mega7: Molecular evolutionary genetics analyses version 7.0 for bigger database. *Molecular Biology and Evolution*, 33(7), 1870–1874.

Kyalo, M., Gichira, A. W., Li, Z. Z., Saina, J. K., Malombe, I., Hu, G. W., & Wang, Q. F. (2018). Characterization and comparative analysis of the complete chloroplast genome of the critically endangered species *Streptocarpus teiensiens* (Gesneriaceae). *Biomedical Research International*, 11, 1057847.

Llibrado, P., & Rozas, J. (2009). DanSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451–1452.

Li, G., Chen, A. J., Chen, X. Y., Li, X. L., & Gao, W. W. (2010). First report of *Amomum volsum* (cardamom) leaf lesion caused by *Pyricularia costina* in China. *New Discovery*, 3(5), 159–163.

Liu, X., Zhou, B. Y., Yang, H. Y., Li, Y., Yang, Q., Lu, Y. Z., & Gao, Y. (2018). Sequencing and characterization of *Chrysanthemum carinatum* Schousb and *Kalimeris indica* L. *European Journal of Horticultural Science*, 92(2), 1–10.

Liu, X., Yang, H., Zhao, B., Li, T., & Xiang, B. (2017). The complete chloroplast genome sequence of the folk medicinal and vegetable plant purslane (*Portulaca oleracea* L.). *International Journal of Biological Sciences*, 19(1), 313–321.

Molecules, 23(4), 857.

Moore, C. W. (1983). Chloroplast genome sequence of *Nuphar advena*. *Molecular and General Genetics*, 194(2), 156–159.

Xu, X., Yang, D., Wang, D., Xu, X., Zhu, L., & Zhao, Z. (2015). Solidification of floating organic drop liquid-phase microextraction cell fishing with gas chromatography-mass spectrometry for screening bioactive components from *Amomum volsum* Lour. *Biomedical Chromatography*, 29(4), 626–632.

Yang, H. R., Xia, J., Zhang, J. E., Yang, J. Z., Zhao, H. H., Wang, G., & Sun, J. J. (2018). Characterization of the complete mitochondrial genome sequences of three Croakers (*Perciformes, Sciaenidae*) and novel insights into the phylogenetics. *International Journal of Molecular Science*, 19(6), 1741.

Yang, M., Zhang, X., Liu, G., Chen, K., Yun, Q., & Zhao, D. (2012). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One*, 7 (9) e12762.

Yi, D. K., Lee, H. L., Sun, B. Y., Chung, M. Y., & Kim, K. J. (2012). The complete chloroplast DNA sequence of *Fistuliflorus serrutus* (Acalyphaceae): Comparative evolutionary analysis with other asteraceae. *Molecules and Cells*, 33(5), 497–508.

Yin, C. K., Zhang, Y., Li, Y. J., & Du, F. K. (2018). Different, natural selection pressures on the atp6 gene in evergreen clorophyllous and deciduous oak species: Evidence from comparative analysis of the complete chloroplast genome of *Quercus aquifolioides* with other oak species. *International Journal of Molecular Science*, 19(4), 1042.

Yu, X. Q., Drew, B. T., Yang, J. B., Gao, L. M., & Lu, D. Z. (2017). Comparative chloroplast genomes of eleven *Schima* (Theaceae) species: Insights into DNA barcoding and phylogeny. *PLoS One*, 12(6) e0178026.

Zhang, G. F., Zhong, Z. M., Huang, S., & Lai, X. P. (2018). Identification and clustering analysis of *Schisandrae* with mtDNA barcode. *Lishizhen Medicine and Materia Medica Research*, 29(1), 99–102.

Zhang, X. G. (1994). Identification of *Amomum volsum* and its confused varieties. *Chinese Herbal Medicines*, 8, 498–501.

Zhang, Y., Su, L., Gao, A., Chen, J., Wu, L., Hu, W., & Zhang, W. (2016). The complete chloroplast genome sequences of five *Epimedium* species: Lights into phylogenetic and taxonomic analyzing. *Frontiers in Plant Science*, 7, 306.

Zhou, J., Cui, Y., Chen, X., Li, Y., Xu, Z., Duan, B., & Li, Y. (2018). Complete chloroplast genomes of *Populus trichocarpa* and *Populus tremuloides*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecular*, 23(2), 437.

Zuo, L., Shang, A., Zhang, S., Yu, X., Ren, Y., Yang, M., & Wang, J. (2017). The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: Genome structure, comparative analysis, and phylogenetic significance. *Trends in Ecology & Evolution*, 12(2), e0171264.