

## RESEARCH ARTICLE

# The complete chloroplast genome sequence of *Alpinia oxyphylla* Miq. and comparison analysis within the Zingiberaceae family

Bingmiao Gao<sup>1,2</sup>, Lin Yuan<sup>2</sup>, Tianle Tang<sup>3</sup>, Jie Hou<sup>2</sup>, Kun Pan<sup>2\*</sup>, Na Wei<sup>1,2\*</sup>

**1** Hainan Provincial Key Laboratory of R&D on Tropical Herbs, Hainan Medical University, Haikou, China, **2** School of Pharmacy, Hainan Medical University, Haikou, China, **3** Environmental Science, School of Tropical and Laboratory Medicine, Hainan Medical University, Haikou, China

\* [pankun001219@163.com](mailto:pankun001219@163.com) (KP); [weina-0613@163.com](mailto:weina-0613@163.com) (NW)



## Abstract

*Alpinia oxyphylla* Miq. (*A. oxyphylla*) is an important edible and traditional herbal medicine. In this study, the complete chloroplast genome of *A. oxyphylla* was sequenced, analysed, and compared to five species in the Zingiberaceae family. The size of the *A. oxyphylla* chloroplast genome was 161351 bp, which consisted of a large single-copy (LSC, 87248 bp) and small single-copy (SSC, 16175 bp) region separated by a pair of inverted repeats (IRA and IRb, 28964 bp each). The genome encoded 132 unique genes, including 87 protein-coding genes, 37 tRNAs and four rRNAs. The GC content of the genome was 36.17%. A total of 53 simple sequence repeats (SSRs) and 80 long repeats were identified in the *A. oxyphylla* chloroplast genome. The chloroplast genome of *A. oxyphylla* shared the highest sequence similarity of >90% with the chloroplast genome of *A. zerumbet*, and six chloroplast genomes in the Zingiberaceae family were compared by using CGView Comparison Tool (CCT). According to the phylogenetic tree, the Zingiberaceae family is divided into two categories, which coincide with the classification of the characteristics of sun-like and shade-like in plants. Our results reveal the phototrophic component of NADH-dehydrogenase (ndhB and ndhC), photosystem II (psbZ) and ATP synthase (atpE, atpF) exhibit adaptive evolution under different environments, and the strength of light is an important trigger for the adaptations at the chloroplast level.

## OPEN ACCESS

**Citation:** Gao B, Yuan L, Tang T, Hou J, Pan K, Wei N (2019) The complete chloroplast genome sequence of *Alpinia oxyphylla* Miq. and comparison analysis within the Zingiberaceae family. PLoS ONE 14(6): e0218817. <https://doi.org/10.1371/journal.pone.0218817>

**Editor:** Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

**Received:** January 14, 2019

**Accepted:** June 10, 2019

**Published:** June 24, 2019

**Copyright:** © 2019 Gao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The complete chloroplast genome of *A. oxyphylla* are available from GenBank accession number: KY985237.

**Funding:** Funded by National Natural Science Foundation of China (No. 81560611), Hainan Provincial Keyjoint Research and Invention Program (No. ZDYF2018138), China Scholarship Council (No. 201708460079), and the Program of Hainan Association for Science and Technology Plans to Youth R & D Innovation (No. 201513).

## Introduction

*Alpinia oxyphylla* Miq. (*A. oxyphylla*) is an important edible and traditional Chinese herbal medicine that originates in Hainan and is widely cultivated in southern China [1]. The fruits of *A. oxyphylla* have been used as valuable medicines that have a long clinical history and are often used as condiment food in China [2,3]. Numerous studies have reported that *A. oxyphylla* is rich in flavonoids, sesquiterpenes, diterpenes, and diarylheptanoids, which have many pharmacological effects, such as improved memory, anti-oxidation, anti-inflammatory, neuroprotective and anticancer [4–8].

**Competing interests:** The authors have declared that no competing interests exist.

Chloroplasts are small organelles inside the cells of plants that contain photosynthetic machinery and produce essential energy for plants [9]. Chloroplasts have their own genetic systems, which consist of a closed circular DNA molecule [9,10]. In recent years, chloroplast genomes have been commonly used for the identification and phyletic evolution analysis of species because of their conserved gene sequences and important role in plants [11]. With the development of high-throughput DNA sequencing technologies, there has been an explosion in the number of available chloroplast genome sequences. However, the chloroplast genome sequences of medicinal plants still require further study. To date, five species of the Zingiberaceae family plant chloroplast genome have been reported, namely, *Alpinia zerumbet*, *Amomum kravanh*, *Curcuma roscoeana*, *Curcuma flaviflora* and *Zingiber spectabile* [12–15]. *A. oxyphylla* chloroplast genome sequences have not yet been reported, which has seriously hindered the development of genetic diversity and breeding of *Alpinia* plants. Therefore, it is highly important and essential to study the phylogeny and evolution of Zingiberaceae plants [16].

In this study, we reported the complete chloroplast genome sequence of *A. oxyphylla*, including a description of its general features, IR contraction and expansion, codon usage and analysis of SSRs and long repeats. In addition, six chloroplast genome sequences in the Zingiberaceae family were compared by using CGView Comparison Tool (CCT). Moreover, we constructed a phylogenetic tree of the Zingiberales, which provides basic genetic information on the genetic diversity and breeding of *Alpinia* plants.

## Materials and methods

### Ethical statement

No specific permits were required for the collection of specimens for this study.

### Plant material, DNA extraction, and sequencing

Fresh *A. oxyphylla* leaves were collected from cultivated fields in Hainan Province, China. Total genomic DNA was extracted from 100 mg of fresh leaves using the Plant Genomic DNA Kit with a standard protocol (Tiangen, Beijing, China). Purified genomic DNA was quantified with a Nanodrop 2000 spectrometer (Thermo Fisher Scientific, Wilmington, USA). Normalized genomic DNA was used to generate a 500 bp (insert size) paired-end library, following the Illumina HiSeq4000 standard protocol. Approximately 2 G raw data were generated with read lengths of 150 bp, and the chloroplast genome sequencing depth was nearly 60×.

### Chloroplast genome assembly and annotation

First, Illumina paired-end reads were filtered on the basis of quality values, and the low-quality reads were trimmed. The remaining clean reads were used for assembly with SOAPdenovo2 (<http://soap.genomics.org.cn/soapdenovo.html>) on the basis of overlapping and paired-end relationships. Next, all clean reads were mapped onto the assembled contigs to obtain a complete chloroplast genome sequence. Genome confirmation was indispensable to perform after assembly. Finally, the paired-end clean reads were mapped onto the assembled genome with 100% coverage, and the insert-size matched the information of the sequenced library.

Annotation was performed using the online program Dual Organellar GenoMe Annotator (DOGMA) [17]. To prove the correctness of gene and exon boundaries, putative gene and protein sequences were BLAST searched in the Nt and Nr databases. The tRNA genes of *A. oxyphylla* were further verified using the online tRNAscan-SE and tRNADB-CE search servers [18–20]. The map of the circular *A. oxyphylla* chloroplast genome was drawn through Organellar Genome DRAW (OGDRAW v1.2) [21].

## Genome structure analyses and genome comparison

The distribution of codon usage was detected by the software CodonW (University of Texas, Houston, TX, USA) with the relative synonymous codon usage (RSCU) ratio [22]. The mVISTA program in Shuffle-LAGAN mode was applied to compare the *A. oxyphylla* chloroplast genome with five other chloroplast genomes. The boundaries between the IR and SC regions of *A. oxyphylla* and five other Zingiberaceae species were compared and analysed. The visualization of codon usage in the form of heatmaps of 17 species of Zingiberales and a histogram were conducted with R language with an RSCU value.

## Repeat sequence analyses

Repeat sequences in chloroplast genomes were detected by the REPuter program [23], including forward, reverse, palindrome, and complement sequences in the chloroplast genome of *A. oxyphylla*. The length and identity of the repeats were limited to  $\geq 30$  bp and  $>90\%$ , respectively [24]. The SSRs were searched using MISA [25], with the following repeat threshold settings: 10 repeats for mono-nucleotide, 5 repeats for di-, 4 repeats for tri-nucleotide, and 3 repeats for tetra- and penta-nucleotide SSRs [26].

## CCT map

The *A. oxyphylla* chloroplast genome was compared with other available chloroplast genomes of Zingiberales by using CCT [27]. Genes were assigned by Clusters of Orthologous Groups, and BLAST was used to align other genomes to that of *A. oxyphylla*. The visualization of the circular map was conducted by CCT. AT distributions were measured on the basis of AT skewed using the equation:  $AT\text{-skew} = (A-T)/(A+T)$ .

## Phylogenetic analysis

Concatenated alignments of 17 chloroplast genome sequences were performed using MUSCLE v.3.8.31. The phylogenetic analysis was carried out using the ML method with RAxML8.1, and the trees were visualized and annotated using the tree viewer of MEGA6 [28]. Statistical supports were assessed with 1000 bootstrap pseudo-replicates.

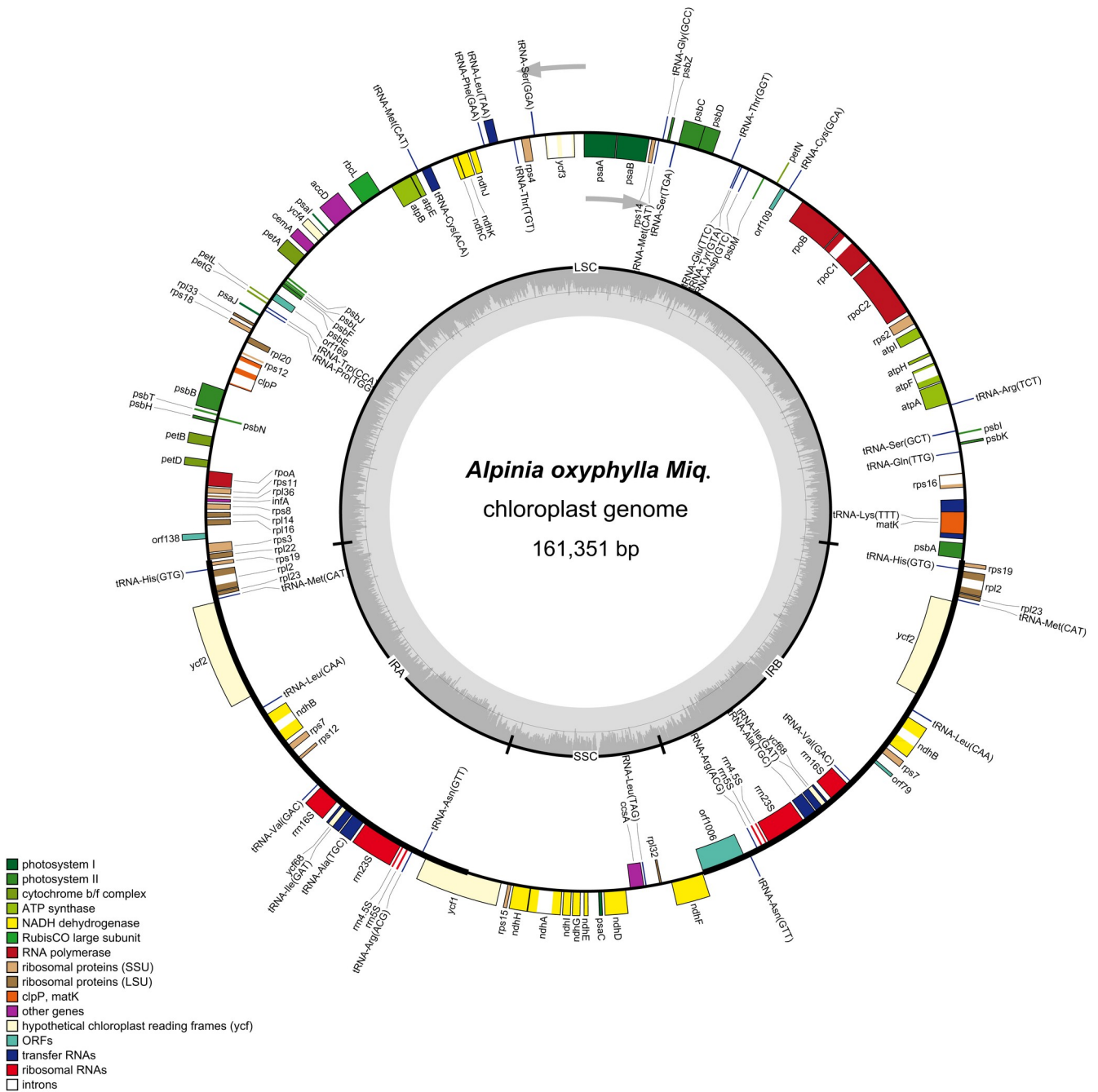
## Positive selection analysis of protein sequence

To investigate the evolutionary process of light adaptation of Zingiberaceae plants, we calculated the Nonsynonymous (Ka), Synonymous (Ks) and Ka/Ks ratios of protein coding genes associated with the photosystem using KaKs\_Calculator 2.0 [29].

## Results and discussion

### General features of the *A. Oxyphylla* chloroplast genome

The complete chloroplast genome of *A. oxyphylla* (GenBank Accession Number: KY985237) has a typical quadripartite structure and is a circular molecule 161,351 bp in size (Fig 1 and Table 1). The genome contains a small single-copy (SSC) region of 16175 bp and a large single-copy (LSC) region of 87248 bp, separated by a pair of inverted repeats (IRa and IRb) of 28964 bp each (Fig 1 and Table 1). The GC content of the *A. oxyphylla* chloroplast genome is 36.17%, which is similar to other chloroplast genomes previously reported [14,15,30]. The genome consists of 132 genes, including 87 distinct protein-coding genes, four distinct rRNA genes and 37 distinct tRNA genes, 21 of which were duplicated in the IR regions, 12 in the SSC region and 84 in the LSC region (Table 1).



**Fig 1. Gene map of the complete chloroplast genome of *A. oxyphylla*.** Genes on the outside of the circle are transcribed clockwise, while those inside are counterclockwise. Genes belonging to different functional groups are colour-coded. The darker and lighter grey in the inner circle correspond to GC and AT content, respectively.

<https://doi.org/10.1371/journal.pone.0218817.g001>

The size of the *A. oxyphylla* chloroplast genome was similar to those of five Zingiberaceae family species (Table 2). The size of the *A. kravanh* chloroplast genome (162766 bp) is the longest, and the *Z. spectabile* chloroplast genome (155890 bp) is the shortest. Interestingly, the SSC region (15,390 bp) of *A. kravanh* is the shortest, whereas the SSC region (18611 bp) of the *Z. spectabile* chloroplast genome is the longest. Five of the chloroplast genomes contain 132

**Table 1. Gene contents in the complete chloroplast genome of *A. oxyphylla*.**

Category of genes	Group of gene	Name of gene
<b>Self-replication</b>	Small subunit of ribosome (SSU)	rps2, rps3, rps4, rps7*, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19*
	Large subunit of ribosome (LSU)	rpl2*, rpl14, rpl16, rpl20, rpl22, rpl23*, rpl32, rpl33, rpl36
	DNA-dependent RNA polymerase	rpoA, rpoB, rpoC1, rpoC2
	Ribosomal RNA genes (rRNA)	rrn16S*, rrn5S*, rrn23S*, rrn4.5S*
	Translation initiation factor	infA
	Transfer RNA genes (tRNA)	trnA-TGC*, trnC-GCA, trnC-ACA, trnD-GTC, trnE-TTC, trnF-GAA, trnG-GCC, trnH-GTG*, trnI-GAT*, trnK-TTT, trnL-TAA, trnL-CAA*, trnL-TAG, trnM-CAT*, trnN-GTT*, trnP-TGG, trnQ-TTG, trnR-TCT, trnR-ACG*, trnS-GCT*, trnS-TGA, trnS-GGA, trnT-GGT, trnT-TGT, trnV-GAC*, trnW-CCA, trnY-GTA
<b>Genes for photosynthesis</b>	Subunits of NADH dehydrogenase	ndhA, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
	Large subunit of Rubisco	rbcL
	Subunits of photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	Subunits of photosystem I	psaA, psaB, psaC, psaI, psaJ
	Subunits of ATP synthase	atpA, atpB, atpE, atpF, atpH, atpI
	Subunits of cytochrome	petA, petB, petD, petG, petL, petN
	Photosystem I assembly	ycf3, ycf4
<b>Other genes</b>	Envelope membrane protein	cemA
	C-type cytochrome synthesis gene	ccsA
	Subunit of acetyl-CoA	accD
	Protease	clpP
	Maturase	matK
<b>Unknown</b>	Conserved open reading frames	orf109, orf138, orf169
<b>Pseudogenes</b>		ycf1, ycf2, ycf68*, orf79, orf1006

Note

\* indicates a duplicated gene.

<https://doi.org/10.1371/journal.pone.0218817.t001>

**Table 2. Comparison of the general features of the six Zingiberaceae chloroplast genomes.**

Genome feature	<i>A. oxyphylla</i>	<i>A. zerumbet</i>	<i>A. kravanh</i>	<i>C. roscoeana</i>	<i>C. flaviflora</i>	<i>Z. spectabile</i>
GenBank	KY985237	JX088668	MF991963.1	KF601574	NC_028729.1	NC_020363.1
Size (bp)	161351	159773	162766	159512	160478	155890
LSC (bp)	87248	87644	87728	87015	88008	85983
SSC (bp)	16175	18295	15390	18531	18570	18611
IR (bp)	28964	26917	29824	26983	26950	25554
Total genes	132	132	135	132	132	132
Protein genes	87	86	80	86	86	86
tRNA genes	37	38	30	38	38	38
rRNA genes	4	4	4	4	4	4
GC (%)	36.17%	36.27%	31.3%	36.33%	36.30%	36.29%

<https://doi.org/10.1371/journal.pone.0218817.t002>

genes, excluding only *A. kravanh* (135 genes). As shown in Table 2, the *C. roscoeana* chloroplast genome has the highest GC content (36.33%), while the *A. kravanh* chloroplast genome has the lowest GC content (31.3%). In addition, 87 protein genes were identified in *A. oxyphylla*, 80 were identified in *A. kravanh*, and 86 were identified in the other four species. Four conserved rRNAs were identified in every species. The *A. oxyphylla* chloroplast genome encodes 37 types of tRNAs, *A. zerumbet*, *C. roscoeana*, *C. flaviflora* and *Z. spectabile* encode 38, whereas *A. kravanh* encodes 30 (Table 2).

### IR contraction and expansion

The contraction and expansion of the IR region are common evolutionary events and are considered the major reasons for size differences in different chloroplast genomes, which is best for studying the phylogeny and the chloroplast genome evolution history of early land plant lineages [31,32]. In the *A. oxyphylla* chloroplast genomes, the boundary of IR/LSC extended into the *rps19* gene, and 129 bp of *rps19* extended into the IR region; the boundary of IR/SSC extended into the *ndhF* gene, and 42 bp of *ndhF* extended into the IR region; the boundary of IRb/SSC extended into 3015 bp of *ycf1*; and the boundary of IRb/LSC and IRa/LSC extended into the *rpl22* and *psbA* genes, respectively. In this study, a detailed comparison of the borders among the IR, LSC and SSC regions among the six Zingiberaceae chloroplast genomes is presented in Fig 2. The pseudogene *ycf1* is often used to study genetic variation in the chloroplast genome in higher plants [32], and the length ranges from 924 to 3888 bp in the six comparable chloroplast genomes. The *ndhF* gene was 32, 37, 42, 136, 136 and 251 bp from the IRb and SSC border in *Z. spectabile*, *A. kravanh*, *A. oxyphylla*, *C. roscoeana*, *C. flaviflora* and *A. zerumbet*, respectively. The *rps19* gene is usually one of the most abundant transcripts among the plant chloroplast genome, including five comparable chloroplasts. However, the *rps19* gene was completely located in the LSC region in the *Z. spectabile* chloroplast genome. Our results

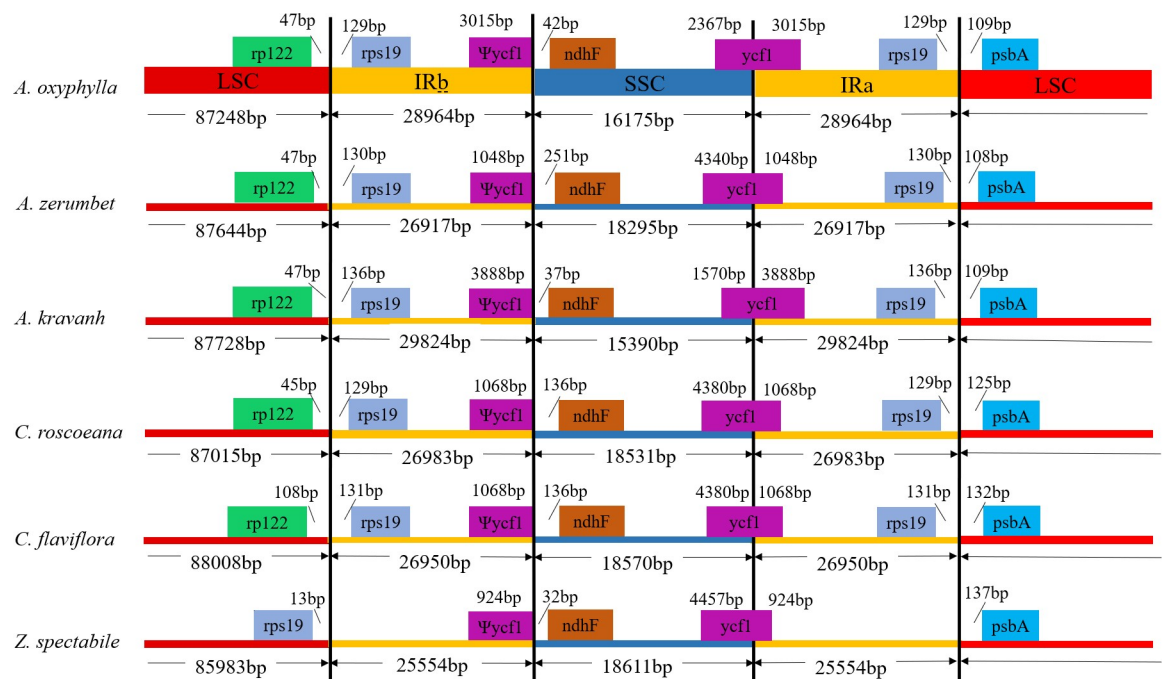


Fig 2. Comparison of the borders of the LSC, SSC and IR regions among the six chloroplast genomes. Ψ, pseudogenes.

<https://doi.org/10.1371/journal.pone.0218817.g002>

suggest that the IR/LSC boundary might be conversed among the chloroplast genomes of closely related family species, but greater diversity also occurs between relatively distantly related family species, such as *Z. spectabile* [33,34].

### Codon usage

The standard ATG codon is typically the start codon for most protein-coding genes. However, ATA and ATC are also used as alternatives to ATG as the initiation codon under certain circumstances [35]. The initiation codon ATG of three genes was replaced among the *A. oxyphylla* chloroplast protein-coding genes, which were ATC for rps12 and orf79 and ATA for rp12(S1 Table). The codon usage frequency and relative synonymous codon usage (RSCU) were analysed based on sequences of 87 distinct protein-coding genes in the *A. oxyphylla* chloroplast genome (Fig 3). The high RSCU value was probably attributed to the function of the amino acid or the structure of the peptide to avoid error in transcription [35]. As shown in Fig 4, the result of the distributions and the visualization of codon usage in the form of heatmaps of 17 species of Zingiberales suggested that approximately one-third of the codons were not frequently used. These codons are shown in blue, which indicates an RSCU value of <1 and weak codon bias. The results showed evident codon use preferences for *A. oxyphylla*, among which AGA, TTA, GCT, TCT, and AGA were used most frequently (Fig 4). Approximately two-thirds of all codons of *A. oxyphylla* that had high RSCU values showed a high A/T preference in the third codon. This phenomenon is common in the chloroplast genomes of higher plants [36].

### Analysis of SSRs and repeats

Simple sequence repeats (SSRs), also known as microsatellites, are a group of tandem repeated sequences, generally ranging in length from 1–6 or more base pairs and are widely distributed in chloroplast genomes [37]. A total of 53 SSRs were detected from the *A. oxyphylla* chloroplast

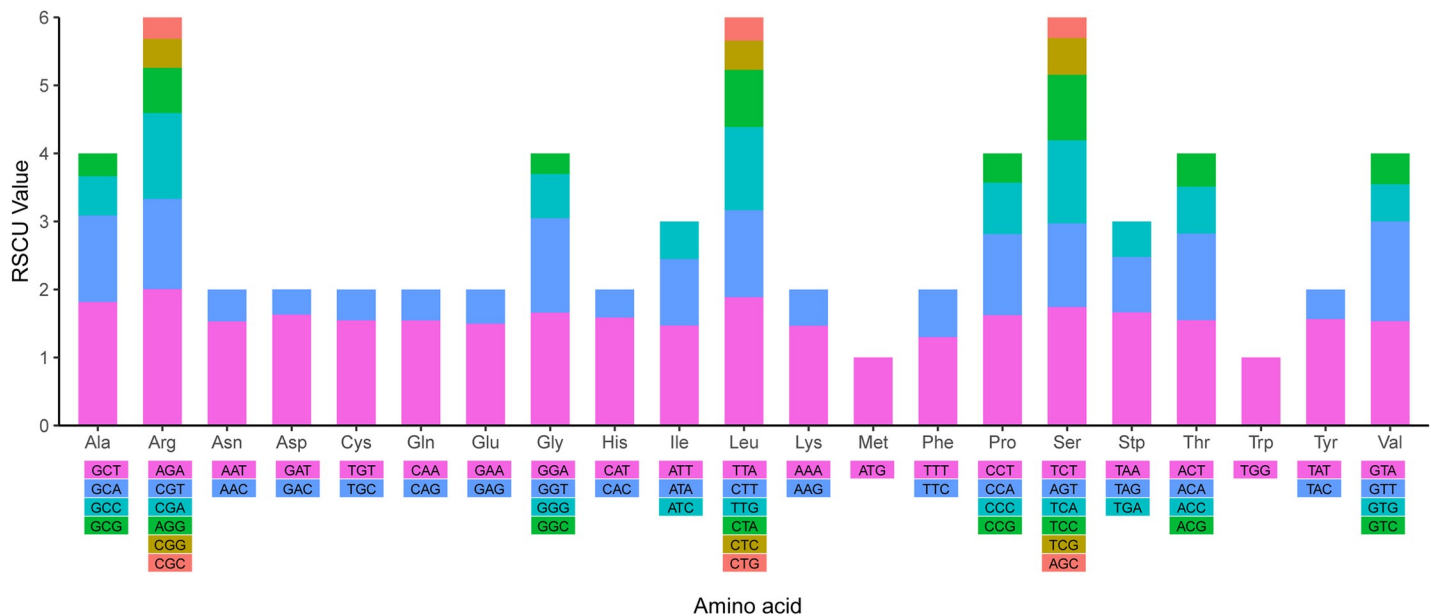
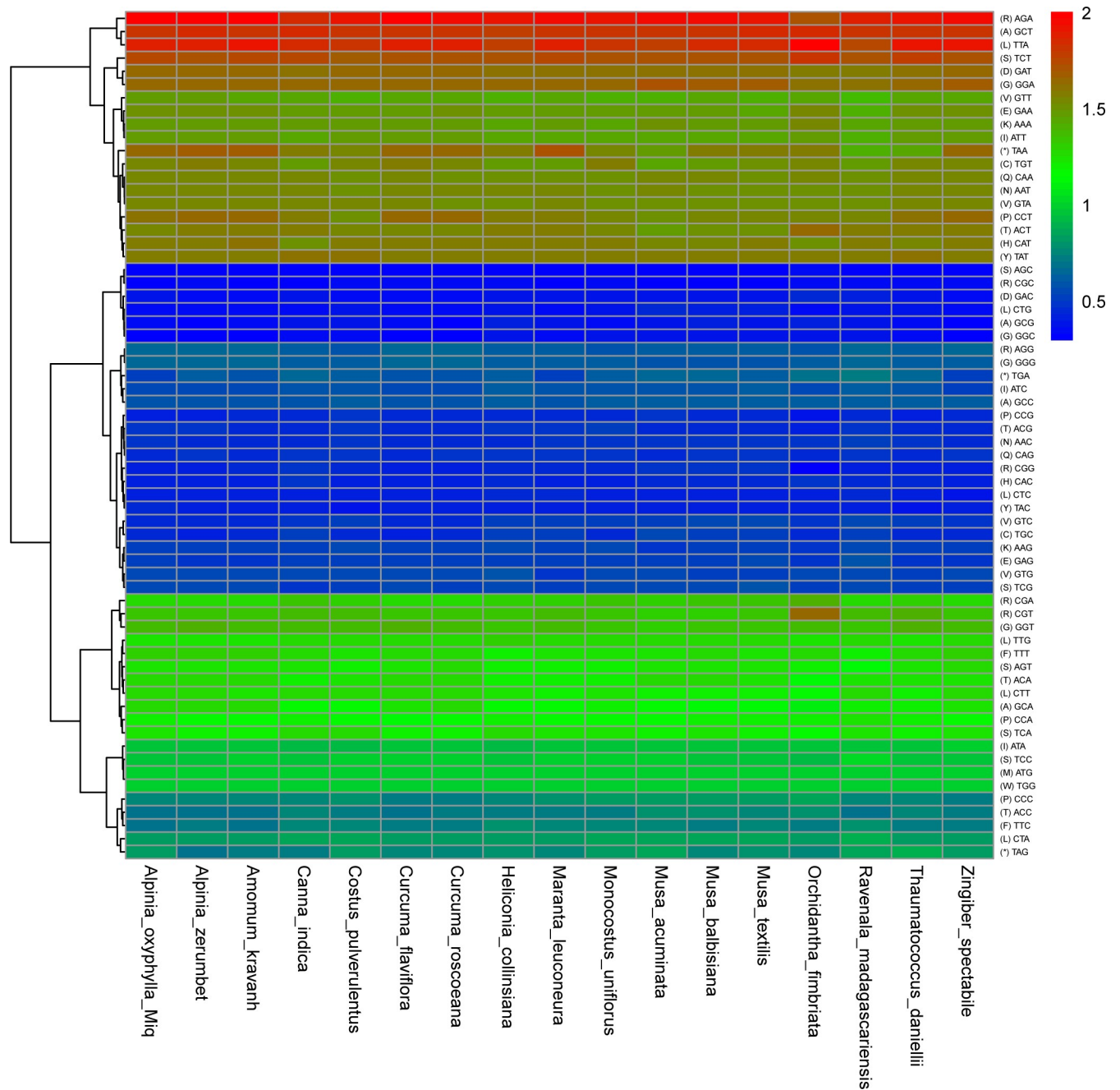


Fig 3. Codon content of 20 amino acids and stop codons in all protein-coding genes of the *A. oxyphylla* chloroplast genome. The colour of the histogram corresponds to the colour of codons.

<https://doi.org/10.1371/journal.pone.0218817.g003>

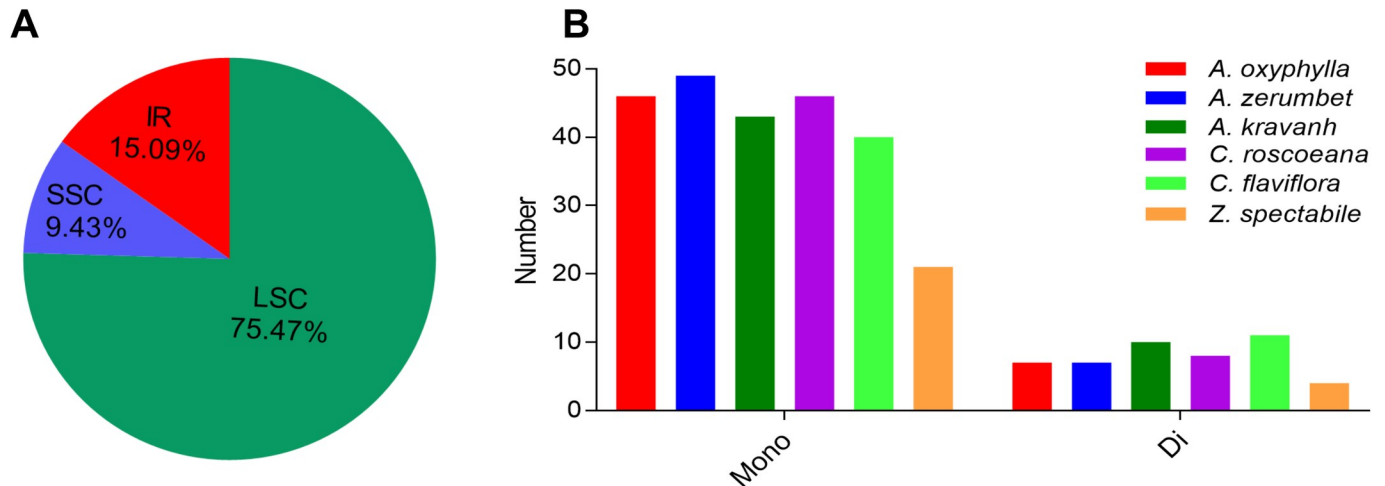


**Fig 4. Heatmap analysis for codon distribution of all protein-coding genes of all considered species.** Colour key: higher red values indicate higher RSCU values, and lower blue values indicate lower RSCU values.

<https://doi.org/10.1371/journal.pone.0218817.g004>

genome, including 46 mono- and 7 di-nucleotide SSRs, which were located in the LSC region (75.47%), IR region (15.09%) and SSC region (9.43%), respectively (Fig 5 and S2 Table). Furthermore, the distribution pattern and number of SSRs among the six Zingiberaceae chloroplast genomes (*A. oxyphylla*, *A. kravanh*, *A. zerumbet*, *C. flaviflora*, *C. roscoeana*, and *Z. spectabile*) were compared, and the results suggested that there was little difference in the distribution pattern and number of SSRs among the six chloroplast genomes (Fig 5). In addition,





**Fig 5. SSR analysis of the six Zingiberaceae chloroplast genomes.** (A) Presence of SSRs in the LSC, SSC and IR regions (*A. oxyphylla*); (B) Presence of polymers in the chloroplast genome of *A. oxyphylla*, *A. zerumbet*, *A. kravanh*, *C. roscoeana*, *C. flaviflora*, and *Z. spectabile*.

<https://doi.org/10.1371/journal.pone.0218817.g005>

a total of 80 repeats were detected in the *A. oxyphylla* chloroplast genome, including 4 complement, 35 forward (direct) repeats, 36 palindrome (inverted) repeats and 14 reverse repeats (S3 Table). Fifty-seven repeats were located in the intergenic spacers (IGS) regions, 19 repeats were located in coding sequence (CDS) regions, and 4 repeats were located in intron regions. These SSRs and repeats can be made into lineage-specific markers, which can provide genetic diversity analysis for *A. oxyphylla* and its related species [38,39].

### CCT map

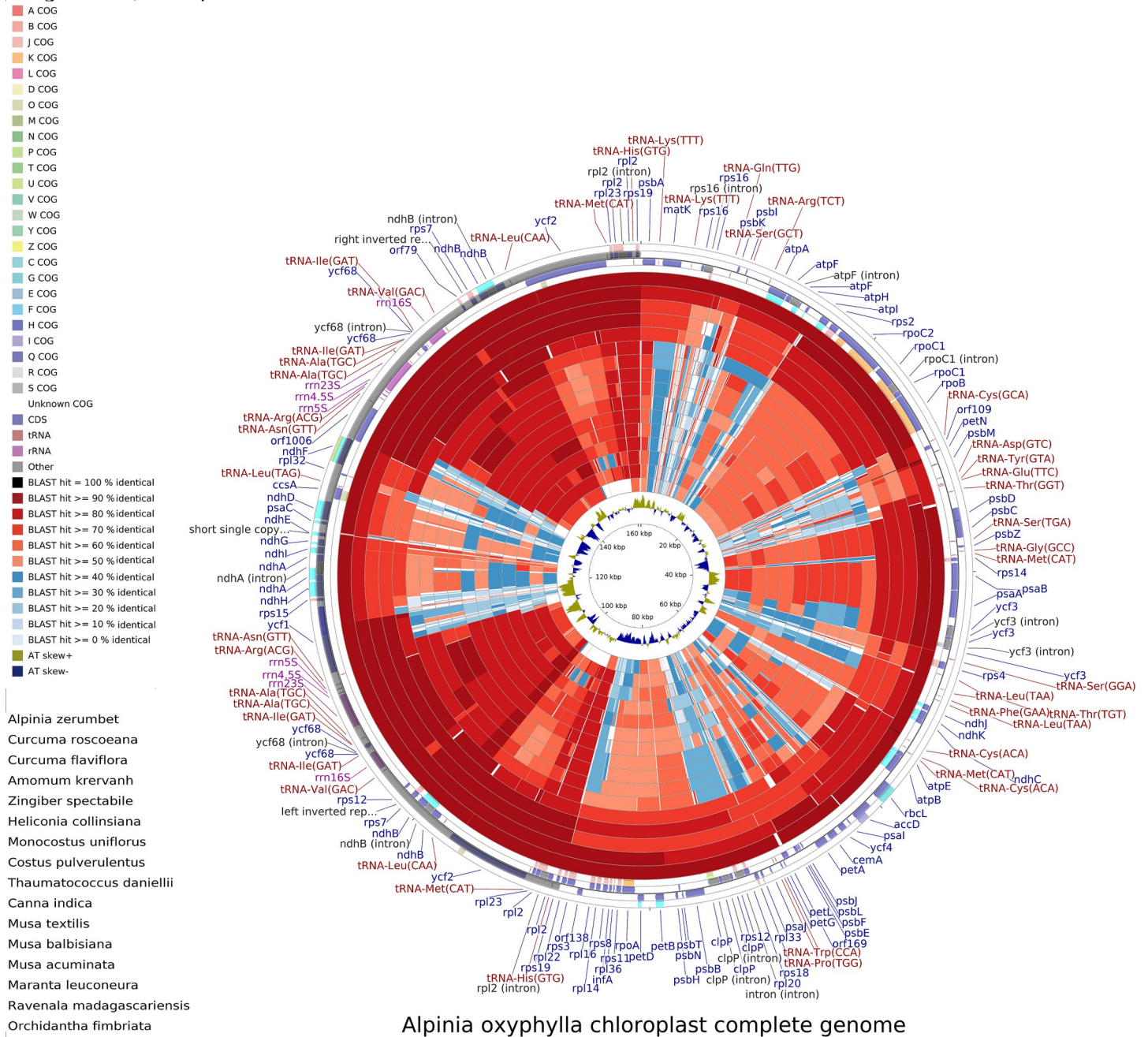
CCT is a package for visually comparing circular bacterial, plasmid, chloroplast, or mitochondrial genome sequences [40]. The *A. oxyphylla* chloroplast genome was compared with 16 previously reported chloroplast genomes of Zingiberales by using CCT (Fig 6). The results showed that the highest sequence similarity (>90%) was between the chloroplast genomes of *A. oxyphylla* and *A. zerumbet*, which was consistent with the result of the phylogenetic analysis (Figs 6 and 7). The most similar region appears in the IR region, and diversity exists in the LSC and SSC regions among 17 chloroplast genomes. This evolutionary feature of the chloroplast genome has also been reported in other plants [41].

### Phylogenetic analysis

Complete chloroplast genomes contain a substantial amount of phylogenetic information, which has been used for phylogenetic analysis of deep relationships among the primary clades of Zingiberales [14,16,34,42]. To identify the evolutionary position of *A. oxyphylla* within Zingiberales, an improved resolution of phylogenetic relationships was achieved by using these complete chloroplast genome sequences of 17 Zingiberales species. The maximum likelihood (ML) bootstrap values had values of 100% bootstrap support for the Zingiberaceae family except the node of *Z. spectabile* and *C. flaviflora* (92%) (Fig 7). The *A. oxyphylla* chloroplast genome was closely related to *A. zerumbet* and *A. kravanh*, which then formed a cluster with *Z. spectabile*, *C. roscoeana*, and *C. flaviflora* with 100% bootstrap supports. In addition, the four ginger families form a well-supported clade within which the families Zingiberaceae and Costaceae, Marantaceae and Cannaceae are sisters. Therefore, the results are expected to be useful in resolving the deeper branches of the phylogenetic tree and will help expand the

Accession: KY985237

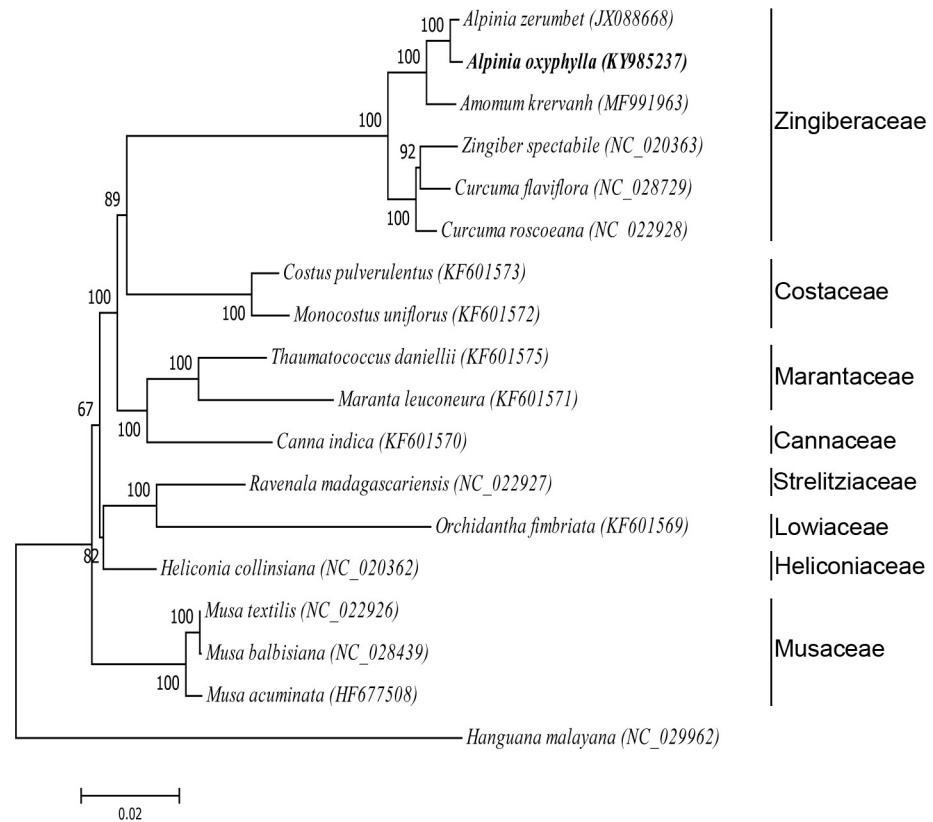
Length: 161,351 bp



**Fig 6. CCT map comparison of 17 chloroplast genomes of *A. oxyphylla* to *Orchidantha fimbriata*.** The four outer rings are the protein-coding gene positions based on the *A. oxyphylla* chloroplast genome. The innermost ring displays AT skew in *A. oxyphylla*. The remaining rings display regions of similarity among the 17 compared chloroplast genomes.

<https://doi.org/10.1371/journal.pone.0218817.g006>

understanding of the evolutionary history of Zingiberaceae, particularly regarding the role of *A. oxyphylla* in plant systematics and evolution.

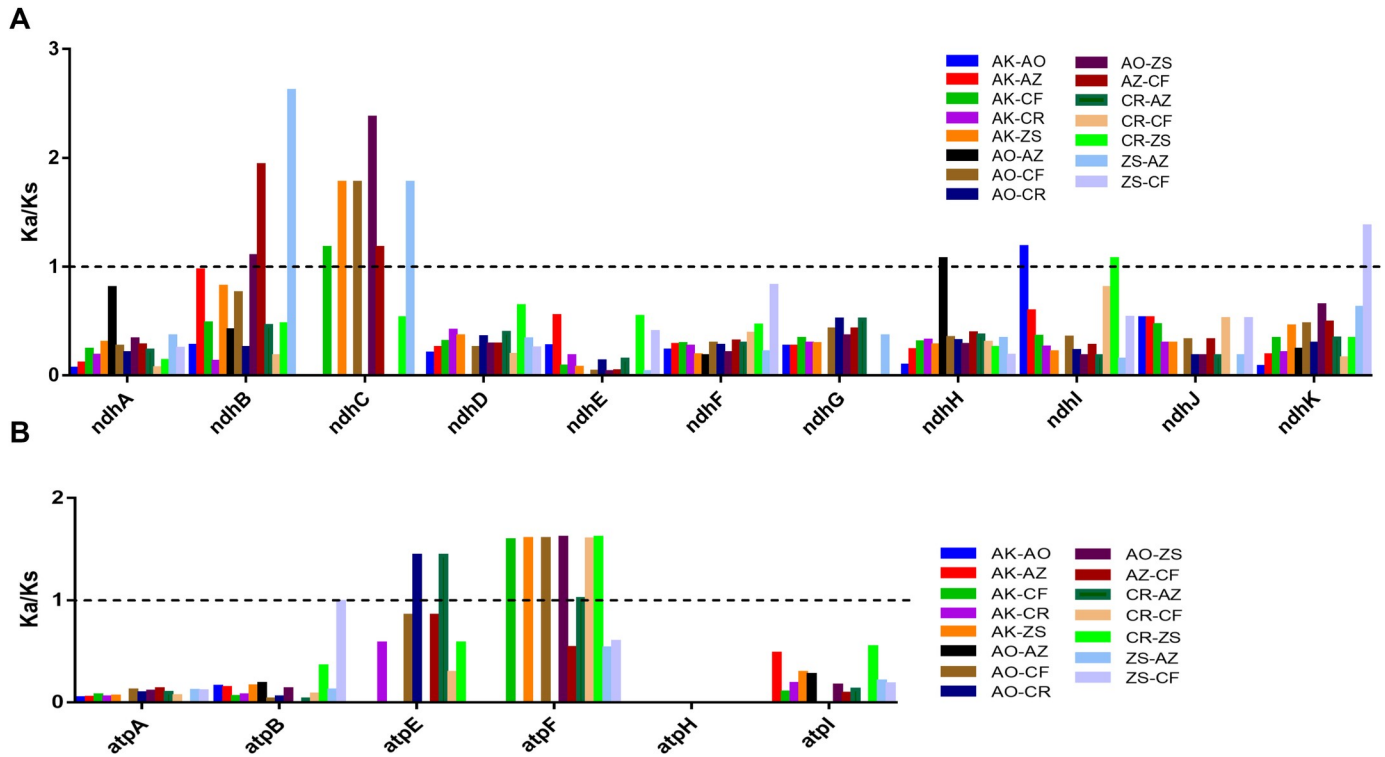


**Fig 7. The maximum parsimony (MP) phylogenetic tree based on chloroplast genome sequences.** The numbers in each node were tested by bootstrap analysis with 1000 replicates.

<https://doi.org/10.1371/journal.pone.0218817.g007>

### Adaptive evolution analysis

The Zingiberaceae family is divided into two main categories, *A. oxyphylla*, *A. kravanh* and *A. zerumbet* clustered into one branch, *C. roscoeana*, *C. flaviflora* and *Z. spectabile* clustered into another (Fig 7). Interesting, the former are typical shade-loving plants whose natural habitats are shade forests of South China, and the latter is sun-loving plants. To investigate the evolutionary process of light adaptation of Zingiberaceae plants, we calculated Ka/Ks ratios of NADH-dehydrogenase, photosystem I, photosystem II, cytochrome b/f complex and ATP synthase coding genes associated with the photosystem. The Ka/Ks is a powerful approach for measuring selective pressure at the protein-coding level. The genes with positive selection played key roles in the adaptation to diverse environment [43]. As a result, NADH-dehydrogenase (*ndhB*, *ndhC*, *ndhH*, *ndhI*, *ndhK*), photosystem II (*psbZ*) and ATP synthase (*atpE*, *atpF*) coding genes with  $Ka/Ks > 1$  were detected, indicating that these genes are undergoing positive selection (Fig 8 and S4 Table). Moreover, the Ka/Ks ratios of the gene *ndhC* and *atpF* in four pairwise comparisons of *A. kravanh*—*C. flaviflora*, *A. kravanh*—*Z. spectabile*, *A. oxyphylla*—*C. flaviflora* and *A. oxyphylla*—*Z. spectabile* were both  $> 1$ , indicating that these two genes are critical in adapting to light (Fig 8). Our results reveal the phototrophic component of NADH-dehydrogenase (*ndhB* and *ndhC*), photosystem II (*psbZ*) and ATP synthase (*atpE*, *atpF*) exhibit adaptive evolution under different environments, and the strength of light is an important trigger for the adaptations at the chloroplast level.



**Fig 8. Ka/Ks ratios of NADH-dehydrogenase (A) and ATP synthase (B) coding genes.** Ka, nonsynonymous; Ks, synonymous; AO, *A. Oxyphylla*; AK, *A. kravanh*; AZ, *A. Zerumbet*; CR, *C. Roscoeana*; CF, *C. Flaviflora*; ZS, *Z. Spectabile*.

<https://doi.org/10.1371/journal.pone.0218817.g008>

### Conclusions

The complete chloroplast sequence of *A. oxyphylla* contains LSC, SSC, and IR regions. A total of 53 SSRs and 80 repeats were identified in the *A. oxyphylla* chloroplast genome. The CCT analytical results indicated that the *A. oxyphylla* chloroplast genome shared the highest sequence similarity with *A. zerumbet*. Phylogenetic trees strongly supported classification of the characteristics of sun-like and shade-like in Zingiberaceae plants. Moreover, the results will help expand the understanding of the evolutionary history of Zingiberaceae, particularly regarding the phototrophic component of NADH-dehydrogenase (*ndhB* and *ndhC*), photosystem II (*psbZ*) and ATP synthase (*atpE*, *atpF*) exhibit adaptive evolution under different environments.

### Supporting information

**S1 Table. Summary of gene/element features.**

(XLS)

**S2 Table. Simple sequence repeats in the *A. oxyphylla* chloroplast genome.**

(XLS)

**S3 Table. Long repeat sequences in the *A. oxyphylla* chloroplast genome.**

(XLS)

**S4 Table. Ka/Ks ratios of protein coding genes associated with the photosystem.**

(XLSX)

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 81560611), Hainan Provincial Keyjoint Research and Invention Program (No. ZDYF2018138), China Scholarship Council (No. 201708460079), and the Program of Hainan Association for Science and Technology Plans to Youth R & D Innovation (No. 201513).

## Author Contributions

**Conceptualization:** Bingmiao Gao.

**Data curation:** Bingmiao Gao, Lin Yuan.

**Formal analysis:** Bingmiao Gao, Tianle Tang, Kun Pan.

**Funding acquisition:** Bingmiao Gao.

**Investigation:** Jie Hou.

**Methodology:** Bingmiao Gao.

**Project administration:** Bingmiao Gao.

**Resources:** Jie Hou.

**Software:** Lin Yuan, Tianle Tang, Jie Hou, Na Wei.

**Supervision:** Kun Pan, Na Wei.

**Writing – original draft:** Bingmiao Gao.

**Writing – review & editing:** Bingmiao Gao, Kun Pan, Na Wei.

## References

1. Liu H, Han CR, Liu HX, Liu YF, He MX. Study on IR fingerprint spectra of *Alpinia oxyphylla* Miq. *Guang Pu Xue Yu Guang Pu Fen Xi*. 2008; 28: 2557–2560. PMID: [19271489](https://pubmed.ncbi.nlm.nih.gov/19271489/)
2. Chang YM, Tamilselvi S, Lin HJ, Tsai CC, Lin YM, Day CH, et al. *Alpinia oxyphylla* Miq extract ameliorates cardiac fibrosis associated with D-galactose induced aging in rats. *Environ Toxicol*. 2018; 1–7.
3. Sharifi-Rad M, Varoni EM, Salehi B, Sharifi-Rad J, Matthews KR, Ayatollahi SA, et al. Plants of the Genus *Zingiber* as a Source of Bioactive Phytochemicals: From Tradition to Pharmacy. *Molecules*. 2017; 22: E2145. <https://doi.org/10.3390/molecules22122145> PMID: [29207520](https://pubmed.ncbi.nlm.nih.gov/29207520/)
4. Chang YM, Tamilselvi S, Lin HJ, Tsai CC, Lin YM, Day CH, et al. *Alpinia oxyphylla* Miq. fruit extract activates IGFR-PI3K/Akt signaling to induce Schwann cell proliferation and sciatic nerve regeneration. *BMC Complement Altern Med*. 2017; 17: 184. <https://doi.org/10.1186/s12906-017-1695-2> PMID: [28359314](https://pubmed.ncbi.nlm.nih.gov/28359314/)
5. Zhang Q, Zheng Y, Hu X, Hu X, Lv W, Lv D, et al. Ethnopharmacological uses, phytochemistry, biological activities, and therapeutic applications of *Alpinia oxyphylla* Miquel: A review. *J Ethnopharmacol*. 2018; 224: 149–168. <https://doi.org/10.1016/j.jep.2018.05.002> PMID: [29738847](https://pubmed.ncbi.nlm.nih.gov/29738847/)
6. Xie Y, Xiao M, Ni Y, Jiang S, Feng G, Sang S, Du G. *Alpinia oxyphylla* Miq. Extract Prevents Diabetes in Mice by Modulating Gut Microbiota. *J Diabetes Res*. 2018; 4230590. <https://doi.org/10.1155/2018/4230590> PMID: [29967794](https://pubmed.ncbi.nlm.nih.gov/29967794/)
7. Yu X, An L, Wang Y, Zhao H, Gao C. Neuroprotective effect of *Alpinia oxyphylla* Miq. fruits against glutamate-induced apoptosis in cortical neurons. *Toxicology letters*. 2003; 144: 205–212. PMID: [12927364](https://pubmed.ncbi.nlm.nih.gov/12927364/)
8. Zhang Q, Hu X, Hui F, Song Q, Cui C, Wang C, Zhao Q. Ethanol extract and its dichloromethane fraction of *Alpinia oxyphylla* Miquel exhibited hepatoprotective effects against CCl<sub>4</sub>-induced oxidative damage in vitro and in vivo with the involvement of Nrf2. *Biomed Pharmacother*. 2017; 91: 812–822. <https://doi.org/10.1016/j.biopha.2017.04.131> PMID: [28501008](https://pubmed.ncbi.nlm.nih.gov/28501008/)
9. Wang X, Zhou T, Bai G, Zhao Y. Complete chloroplast genome sequence of *Fagopyrum dibotrys*: genome features, comparative analysis and phylogenetic relationships. *Scientific reports*. 2018; 8: 12379. <https://doi.org/10.1038/s41598-018-30398-6> PMID: [30120274](https://pubmed.ncbi.nlm.nih.gov/30120274/)

10. Tian N, Han L, Chen C, Wang Z. The complete chloroplast genome sequence of *Epipremnum aureum* and its comparative analysis among eight Araceae species. *PLoS ONE*. 2018; 13: e0192956. <https://doi.org/10.1371/journal.pone.0192956> PMID: 29529038
11. Zhang R, Li Q, Gao J, Qu M, Ding P. The complete chloroplast genome sequence of the medicinal plant *Morinda officinalis* (Rubiaceae), an endemic to China. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016; 27: 4324–4325. <https://doi.org/10.3109/19401736.2015.1089484> PMID: 26465065
12. Zhang Y, Deng J, Li Y, Gao G, Ding C, Zhang L, et al. The complete chloroplast genome sequence of *Curcuma flaviflora* (Curcuma). *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016; 27:3644–3645. <https://doi.org/10.3109/19401736.2015.1079836> PMID: 26367332
13. Wu M, Li Q, Hu Z, Li X, Chen S. The Complete *Amomum kravanh* Chloroplast Genome Sequence and Phylogenetic Analysis of the Commelinids. *Molecules*. 2017; 22: E1875. <https://doi.org/10.3390/molecules22111875> PMID: 29104233
14. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, et al. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? *Ann Bot-London*. 2014; 113: 119–133.
15. Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics*. 2013, 29: 65–87.
16. Gevu KV, Lima HRP, Kress J, Da Cunha M. Morphological analysis of vessel elements for systematic study of three Zingiberaceae tribes. *Journal of Plant Research*. 2017; 130: 527–538. <https://doi.org/10.1007/s10265-017-0911-y> PMID: 28255816
17. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20: 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352> PMID: 15180927
18. Schattner P., Brooks A. N. & Lowe T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*. 2005; 33: W686–W689. <https://doi.org/10.1093/nar/gki366> PMID: 15980563
19. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Research*, 2003; 31: 439–441. <https://doi.org/10.1093/nar/gkg006> PMID: 12520045
20. Abe T, Ikemura T, Sugahara J, et al. tRNADB-CE 2011: tRNA gene database curated manually by experts. *Nucleic Acids Research*. 2011; 39: D210–D213. <https://doi.org/10.1093/nar/gkq1007> PMID: 21071414
21. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*. 2007; 52: 267–274. <https://doi.org/10.1007/s00294-007-0161-y> PMID: 17957369
22. Sharp PM, Li WH. The Codon Adaptation Index—a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications. *Nucleic Acids Research*. 1987; 15: 1281–1295. <https://doi.org/10.1093/nar/15.3.1281> PMID: 3547335
23. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001; 29: 4633–4642. <https://doi.org/10.1093/nar/29.22.4633> PMID: 11713313
24. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999; 27: 573–580. <https://doi.org/10.1093/nar/27.2.573> PMID: 9862982
25. Yang XM, Sun JT, Xue XF, Zhu WC, Hong XY. Development and characterization of 18 novel EST-SSRs from the western flower Thrips, *Frankliniella occidentalis* (Pergande). *Int J Mol Sci*. 2012; 13: 2863–2876. <https://doi.org/10.3390/ijms13032863> PMID: 22489130
26. Li Q, Wan JM. SSRHunter: development of a local searching software for SSR sites. *Yi Chuan*. 2005; 27: 808–810. PMID: 16257914
27. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res*. 2008, 36: W181–184. <https://doi.org/10.1093/nar/gkn179> PMID: 18411202
28. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22: 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
29. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs\_Calculator 2.0: A toolkit incorporating Gamma-Series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics*. 2010; 8: 77–80.
30. Thomas GE, Geetha KA, Augustine L, Mamiyil S, Thomas G. Analyses between Reproductive Behavior, Genetic Diversity and Pythium Responsiveness in *Zingiber* spp. Reveal an Adaptive Significance for Hemiclonality. *Frontiers in Plant Science*. 2016; 7: 1913. <https://doi.org/10.3389/fpls.2016.01913> PMID: 28066470

31. Gu CH, Tembrock LR, Zheng SY, Wu ZQ. The Complete Chloroplast Genome of *Catha edulis*: A Comparative Analysis of Genome Features with Related Species. *International Journal of Molecular Sciences*. 2018; 19: 525.
32. Wu ML, Li Q, Xu J, Li XW. Complete chloroplast genome of the medicinal plant *Amomum compactum*: gene organization, comparative analysis, and phylogenetic relationships within Zingiberales. *Chin Med-Uk*. 2018; 13: 10.
33. Park E, Caplan JL, Dinesh-Kumar SP. Dynamic coordination of plastid morphological change by cytoskeleton for chloroplast-nucleus communication during plant immune responses. *Plant Signaling & Behavior*. 2018; 13: e1500064.
34. Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A. The Complete Chloroplast Genome of Banana (*Musa acuminata*, Zingiberales): Insight into Plastid Monocotyledon Evolution. *PLoS ONE*. 2013; 8: e67350. <https://doi.org/10.1371/journal.pone.0067350> PMID: 23840670
35. Li Y, Kuang XJ, Zhu XX, Zhu YJ, Sun C. Codon usage bias of *Catharanthus roseus*. *Zhongguo Zhong Yao Za Zhi*. 2016; 41: 4165–4168. <https://doi.org/10.4268/cjcm20162213> PMID: 28933083
36. Gichira AW, Li ZZ, Saina JK, Long Z, Hu G, Gituru RW, et al. The complete chloroplast genome sequence of an endemic monotypic genus *Hagenia* (Rosaceae): structural comparative analysis, gene content and micro satellite detection. *PeerJ*. 2017; 5: e2846. <https://doi.org/10.7717/peerj.2846> PMID: 28097059
37. Meng J, Li X, Li H, Yang J, Wang H, He J. Comparative Analysis of the Complete Chloroplast Genomes of Four *Aconitum* Medicinal Species. *Molecules*. 2018; 23: E1015. <https://doi.org/10.3390/molecules23051015> PMID: 29701675
38. Liu W, Kong H, Zhou J, Fritsch PW, Hao G, Gong W. Complete Chloroplast Genome of *Cercis chuniana* (Fabaceae) with Structural and Genetic Comparison to Six Species in Caesalpinioideae. *Int J Mol Sci*. 2018; 19: E1286. <https://doi.org/10.3390/ijms19051286> PMID: 29693617
39. Li W, Liu Y, Yang Y, Xie X, Lu Y, Yang Z, et al. Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*. *BMC Plant Biol*. 2018; 18: 210. <https://doi.org/10.1186/s12870-018-1421-3> PMID: 30257644
40. Grant JR, Arantes AS, Stothard P. Comparing thousands of circular genomes using the CGView Comparison Tool. *Bmc Genomics*. 2012; 13: 202. <https://doi.org/10.1186/1471-2164-13-202> PMID: 22621371
41. Wang Y, Zhan DF, Jia X, Mei WL, Dai HF, Chen XT, Peng SQ. Complete Chloroplast Genome Sequence of *Aquilaria sinensis* (Lour.) Gilg and Evolution Analysis within the Malvales Order. *Frontiers in Plant Science*. 2016; 7: 280. <https://doi.org/10.3389/fpls.2016.00280> PMID: 27014304
42. Kress WJ, Liu AZ, Newman M, Li QJ. The molecular phylogeny of *Alpinia* (Zingiberaceae): A complex and polyphyletic genus of gingers. *American Journal of Botany*. 2005; 92: 167–178. <https://doi.org/10.3732/ajb.92.1.167> PMID: 21652395
43. Gao C, Deng Y, Wang J. The complete chloroplast genomes of *Echinacanthus* species (Acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. *Front. Plant Sci*. 2019; 9:1989. <https://doi.org/10.3389/fpls.2018.01989> PMID: 30687376