



# Complete chloroplast genome of *Boesenbergia rotunda* and a comparative analysis with members of the family Zingiberaceae

Yvonne Jing Mei Liew<sup>1,2</sup> · Kah-Ooi Chua<sup>3</sup> · Hoi-Sen Yong<sup>4</sup> · Sze-Looi Song<sup>5</sup> · Kok-Gan Chan<sup>4,6,7</sup>

Received: 10 June 2022 / Revised: 15 September 2022 / Accepted: 16 October 2022 / Published online: 26 October 2022  
© The Author(s), under exclusive licence to Botanical Society of Sao Paulo 2022

## Abstract

*Boesenbergia rotunda* (L.) Mansf. is a medically important ginger species of the family Zingiberaceae but its genomic information on molecular phylogeny and identification is scarce. In this work, the chloroplast genome of *B. rotunda* was sequenced, characterized and compared to the other Zingiberaceae species to provide chloroplast genetic resources and to determine its phylogenetic position in the family. The chloroplast genome of *B. rotunda* was 163,817 bp in length and consisted of a large single-copy (LSC) region of 88,302 bp, a small single-copy (SSC) region of 16,023 bp and a pair of inverted repeats (IRA and IRB) of 29,746 bp each. The chloroplast genome contained 113 unique genes, including 79 protein-coding genes, 30 transfer RNA (tRNA) genes and four ribosomal RNA (rRNA) genes. Several genes had atypical start codons, while most amino acids exhibited biased usage of synonymous codons. Comparative analyses with various chloroplast genomes of Zingiberaceae taxa revealed several highly variable regions (*psbK-psbI*, *trnT-GGU-psbD*, *rbcL-accD*, *ndhF-rpl32*, and *ycf1*) in the LSC and SSC regions in the chloroplast genome of *B. rotunda* that could be utilized as molecular markers for DNA barcoding and species delimitation. Phylogenetic analyses based on shared protein-coding genes revealed that *B. rotunda* formed a distinct lineage with *B. kingii* Mood & L.M.Prince, in a subclade that also contained the genera *Kaempferia* and *Zingiber*. These findings constitute the first chloroplast genome information of *B. rotunda* that could be a reference for phylogenetic analysis and identification of genus *Boesenbergia* within the Zingiberaceae family.

**Keywords** Ginger · Illumina sequencing · Nanopore sequencing · Plastid genome · Subfamily Zingiberoideae

## 1 Introduction

*Boesenbergia rotunda* (L.) Mansf., also known as finger root, is a medically important ginger species that belongs to the family Zingiberaceae. It is widely distributed across Southeast Asia, India, Southern China and Sri Lanka (Hooker 1875; Larsen 1996; Larsen et al. 1999; Chen and Xia 2019). Many bioactive metabolites have been isolated from *B. rotunda*, and most of these compounds have shown some medicinal properties such as anti-inflammatory, anti-oxidant, antibacterial, anticancer, and antiviral (Tan et al. 2012; Atun et al. 2018; Mohan et al. 2020; Break et al. 2021). A recent study has also reported that the bioactive compound Panduratin A found in *B. rotunda* extract exhibited potent anti-severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) activity (Kanjanasirirak et al. 2020). Due to these therapeutic properties, certain metabolic pathways in *B. rotunda* have been studied using molecular and metabolic engineering approaches to increase the production of the bioactive compounds (Liew

✉ Kah-Ooi Chua  
kahooi@um.edu.my

✉ Kok-Gan Chan  
kokgan@um.edu.my

<sup>1</sup> University of Malaya Centre for Proteomics Research, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>2</sup> Deputy Vice Chancellor's Office (Research and Innovation), Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>3</sup> Centre for Research in Biotechnology for Agriculture, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>4</sup> Institute of Biological Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>5</sup> Institute for Advanced Studies, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>6</sup> International Genome Centre, Jiangsu University, Zhenjiang, China

<sup>7</sup> Institute of Marine Sciences, Shantou University, Shantou 515063, China

et al. 2020, 2021). However, to date, the genomic information of this ginger is scarce. Identification of *Boesenbergia* species based on morphological observation is challenging due to the low morphological variation of the vegetative parts among Zingiberaceae family (Branney 2005; Li et al. 2019). Therefore, a more comprehensive identification approach that couples molecular analysis of chloroplast genome with morphological characterization and phylogenetic analysis is necessary for accurate identification of the ginger species (Gao et al. 2019; Li et al. 2020).

In angiosperm plants, the photosynthetic organelle, chloroplast provides requisite energy for metabolism of plant cells (Green 2011). Chloroplast genomes have stable structure, genes content and order, and they have low mutation rate as the genetic information is maternally inherited (Daniell et al. 2016). These characteristics make chloroplast genomes ideal for plant species identification and phylogenetic analysis to resolve their evolutionary relationships (Sudianto et al. 2019). In general, the chloroplast genomes are characterized by a circular quadripartite structure that is about 120–165 kb in length, and consisting of a large single copy (LSC), a small single copy (SSC) and a pair of inverted repeats (IR) (Wicke et al. 2011).

The recent development of high-throughput DNA sequencing technologies such as Oxford Nanopore has enabled rapid and cost-effective sequencing of long DNA molecules with relatively simple library prep (Tyson et al. 2018). The platform is increasingly used in sequencing of chloroplast genomes as it confers advantages in generating longer contigs with fewer unresolved gaps for better coverage of long repetitive sequences (Kono and Arakawa 2019). However, the drawback of the long read sequencer is its relatively higher error rate compared to short read sequencing. Therefore, Illumina sequencing that provides short read data is often used in combination for a high-accuracy DNA sequencing (Sun et al. 2021).

In this study, the complete chloroplast genome of *B. rotunda* was determined using Illumina and Nanopore sequencing platforms. We also analyzed various characteristics including the amino acid and codon usage as well as the type and distribution of simple sequence repeats (SSRs) and long repeats in the chloroplast genome. The objectives of this study are also to perform comparative analysis to identify highly variable regions that could serve as molecular markers and to perform a robust phylogenetic analysis to elucidate the taxonomic position of *B. rotunda* in the family Zingiberaceae. Overall, this study characterizes the first complete chloroplast genome sequence of *B. rotunda*, which will serve as an important reference for future species identification and the study on genetic diversity in the family Zingiberaceae.

## 2 Materials and methods

**Plant material and DNA extraction** – *Boesenbergia rotunda* plant was obtained from Kuantan, Pahang, Malaysia (3°50'21.8"N, 103°20'30.3"E). The plant specimen was taxonomically identified, and voucher specimens (voucher number: KLU50137) were deposited in the Rimba Ilmu herbarium, Institute of Biological Sciences, Universiti Malaya. Leaves were collected from three different sites of the *B. rotunda* plant and cut into small pieces. The genomic DNA extraction was performed using the high-salt cetyltrimethylammonium bromide (CTAB) method (Inglis et al. 2018). Briefly, 150 mg of the leaf tissues resuspended in sorbitol wash buffer, and homogenized using TACO Prep Bead Beater system (GeneReach Biotechnology Corp, Taiwan). The homogenized tissues were collected by centrifugation at 3000 rpm for 5 min, washed with sorbitol wash buffer and centrifuged once again. The leaf tissues were then resuspended in high-salt CTAB containing 3 M NaCl, 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 3% (w v<sup>-1</sup>) CTAB, and 1% (w v<sup>-1</sup>) PVP, followed by one hour incubation at 65 °C. Chloroform was then added into the homogenate to ensure complete protein precipitation, and homogenate was centrifuged at 3000 rpm for 15 min. The supernatant was transferred to a new tube containing isopropanol (1 × vol) and 10 µl of homebrewed SPRI beads for gDNA precipitation. After incubation at room temperature for 10 min, the tube was placed on a magnetic rack, and the beads were washed twice with 75% v v<sup>-1</sup> ethanol. The beads were resuspended in 100-µL TE buffer and incubated at 37 °C to facilitate the release of DNA from beads into the buffer. The extracted DNA was electrophoresed on a 1% w v<sup>-1</sup> agarose gel to assess its integrity, and the concentration was measured using DeNovix dsDNA High Sensitivity Kit (DeNovix, USA). The final concentration of genomic DNA obtained was between 30 and 40 ng µL<sup>-1</sup>.

**Chloroplast genome sequencing using the Illumina and Nanopore platforms** – A total of 500 ng of the genomic DNA of *B. rotunda* was measured and fragmented into 350 bp using a Covaris Ultrasonicator (Covaris, Woburn, MA, USA). The fragmented DNA was subsequently prepared into an Illumina sequencing library using NEB Ultra II Illumina library preparation kit (New England Biolabs, USA). DNA sequencing was performed on the Illumina NovaSeq 6000 with 2 × 150 bp paired-end read configuration.

For Nanopore DNA sequencing, the genomic DNA (5 µg) of *B. rotunda* was size-selected using 0.15 × vol of SPRI beads in MgCl<sub>2</sub>-PEG8000 size selection buffer (500 mM MgCl<sub>2</sub>, 5% (v v<sup>-1</sup>) PEG-8000) (Stortchevoi et al. 2020). After the removal of supernatant containing

unbound short DNA fragments, the beads were washed with 75% (v/v<sup>-1</sup>) ethanol and the size-selected DNA was eluted in 30 µL of TE buffer. Then, 2 µg genomic DNA was prepared into an Oxford Nanopore sequencing library using ligation sequencing kit (SQK-LSK109) and the native barcoding expansion 1–12 kit (EXP-NBD104) with reference to the manufacturer's protocol. The library was loaded into a Flongle flow cell and sequenced on a MinION Nanopore sequencer following the procedure for Oxford Nanopore Sequencing Technologies.

**Chloroplast genome assembly and annotation** – Base-calling of the fast5 file generated by MinION was performed using Guppy v4.2.2 (high accuracy mode) to generate the Nanopore sequencing reads in fastQ format. The Illumina raw sequence data were quality-filtered using Trimmomatic v.0.39 to eliminate the low quality and ambiguous reads (Bolger et al. 2014). A hybrid assembly of chloroplast genome was performed with the long Nanopore reads and short Illumina reads using SPAdes-3.13.0 with multi-mer approach (Antipov et al. 2016; Prjibelski et al. 2020). The chloroplast contig was identified from the assembled sequences by a BLAST search against the NCBI standard databases with nucleotide collection. Contigs that showed a match with other chloroplast genomes were further examined for length and gene contents to determine the chloroplast genome contig. The chloroplast genome was examined for terminal repeats using contiguity to evaluate their circularity and completeness (Sullivan et al. 2015). The complete chloroplast genome sequence of *B. rotunda* was deposited in the GenBank under the accession number MZ411538.

Chloroplast genome annotation was performed using GeSeq (Tillich et al. 2017) implemented in CHLOROBOX (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>). *Zingiber officinale* Rosc. (NC\_044775) and *Kaempferia elegans* (Wall.) Baker (NC\_040852) in the NCBI Reference Sequence Database (RefSeq) were used as BLAT reference chloroplast genomes. The tRNAs were annotated using tRNAscan-SE v2.0.7 implemented in CHLOROBOX (Chan and Lowe 2019). The annotated genes were aligned with similar genes of closely related ginger species to review the length and boundaries of each gene.

**Chloroplast genome structure and sequence analyses** – A circular map of the chloroplast genome was generated with OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/cite-OGDraw.html>) (Greiner et al. 2019). The nucleotide composition of the chloroplast genome, amino acid frequency and relative synonymous codon usage value (RSCU) of each protein-coding gene was calculated in MEGA X (Kumar et al. 2018). The simple sequence repeats (SSRs) in the chloroplast genome were identified by the web-based Microsatellite Repeats Finder ([http://insilico.ehu.es/mini\\_tools/microsatellites/](http://insilico.ehu.es/mini_tools/microsatellites/), accessed May 12, 2021). The sequence repeat length was set at 2 to 6, while the minimum number

and length of repeat were set at 3 and 10, respectively. The location and size of various long repeats (forward, reverse, palindromic and complement, all  $\geq 30$  bp) were determined by the software REPuter with the setting of hamming distance and minimal repeat size at 3 and 30 bp, respectively (Kurtz et al. 2001).

**Comparative analyses of chloroplast genomes** – Thirteen available chloroplast genomes of closely related Zingiberaceae ginger species were included in the comparative analyses (Table S1). Multiple chloroplast genomes alignment was performed using MAFFT v7.453 (Katoh and Standley 2013). Using *B. rotunda* as the reference, the chloroplast genome architecture of these ginger species was compared in a sequence variation map generated using the software mVISTA (Frazer et al. 2004). The nucleotide variability ( $\pi$ ) throughout the positions in these chloroplast genomes was calculated using the software DnaSP6.0 (window length set at 800 bp, step size set at 200 bp) (Rozas et al. 2017).

**Phylogenetic analyses** – Zingiberaceae species with available chloroplast genomes in the GenBank at the time of analysis (Table S1) were included in the phylogenetic analysis. The chloroplast genome of *Musa acuminata* Colla HF677508 from family Musaceae was selected as outgroup. The taxon *Musa acuminata* is the type species of the genus *Musa* which is also the type genus of *Musaceae*, a closely related family of Zingiberaceae in the same order Zingiberales. A previous study had shown that Zingiberaceae phylogeny rooted using *Musa* taxa was effective in elucidating the evolutionary relationship between the Zingiberaceae taxa (Wang et al. 2021). The nucleotide sequences of 75 unique protein coding genes identified in every chloroplast genome were acquired for analysis. These DNA sequences were aligned by MAFFT v7.453 (Katoh and Standley 2013). The gaps and the poorly aligned sequence regions revealed in the alignment were removed with trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009), before concatenating the individual genes. The best-fit nucleotide substitution models for each of the genes were determined using ModelFinder based on the Bayesian information criterion (Kalyaanamoorthy et al. 2017). A phylogenetic analysis using the maximum likelihood (ML) method under ultrafast bootstrap algorithm with 10,000 replicates was performed using the IQ-TREE (Nguyen et al. 2015). The MEGA X software was used to visualize the constructed phylogenetic tree (Kumar et al. 2018).

Bayesian analysis was performed with Mr. Bayes v.3.1.2 using the Markov chain Monte Carlo (MCMC) method (Huelsenbeck and Ronquist 2001) with two independent runs of  $2 \times 10^6$  generations with four chains, and with trees sampled every 200th generation. The best-fit nucleotide substitution models were analyzed by Kakusan v.3 (Tanabe 2007), using the Bayesian Information Criterion (Schwarz 1978). Likelihood values for all post-analysis trees and

parameters were evaluated for convergence and burn-in using the “sump” command in MrBayes and the software Tracer v.1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>). The first 200 trees from each run were discarded as burn-in (where the likelihood values were stabilized prior to the burn-in), and the remaining trees were used for the construction of a 50% majority-rule consensus tree. The phylogenetic tree generated was viewed in MEGA X software (Kumar et al. 2018).

### 3 Results

**General characteristics of the *B. rotunda* chloroplast genome** – A total of 6,451,016 Illumina paired-end reads (total bases: 967.7 Mbp) and 37,560 Nanopore long reads (total bases: 263 Mbp, mean read length: 7,014 bp, read length N50: 10,020 bp) were generated using two different sequencing platforms. After quality-filtering, 6,270,924 high-quality Illumina paired-end reads and all Nanopore long reads basecalled using Guppy v4.2.2 in high accuracy mode were used for chloroplast genome assembly. The total length of the *B. rotunda* chloroplast genome was 163,817 bp, and had a 83× coverage. It possessed a circular quadripartite structure and contained four regions, namely a large single copy (LSC) of 88,302 bp and a small single copy (SSC) of 16,023 bp that were separated by two inverted repeat (IR) regions (IRA and IRB) of 29,746 bp each (Table 1, Fig. 1). Although the overall GC content of the chloroplast genome was 36.0%, the GC contents across different regions of the chloroplast genome varied significantly from the highest at 41.1% in the LSC and the IRA regions, to 33.9% in the IRB and 29.3% in the SSC regions (Table 1).

A total of 133 genes were annotated in the chloroplast genome (Table S2). Of these, 113 genes (79 protein-coding genes, 30 tRNA genes and four rRNA genes) were unique in the chloroplast genome (Table 2). All the 20 duplicated genes were located in the inverted repeat regions. Six tRNA genes and nine protein-coding genes possessed an

intron region, while the protein-coding genes *clpP1*, *ycf3* and *rps12* consisted of two intron regions (Table S2). The smallest intron (533 bp) occurred in the gene *trnL-UAA*, while the largest intron (2659 bp) occurred in the gene *trnK-UUU* with its intron region encompassing the gene *matK* in LSC region (Fig. 1, Table S2). Notably, only the gene *rps12* was trans-spliced into two fragments situated in the LSC and the IR regions (Fig. 1). On the other hand, the tRNAs in the chloroplast genome ranged from 70 to 91 bp in length (Table S2). Most of them showed a canonical clover leaf-like secondary structure (Fig. 2a). All the tRNAs possessed an acceptor arm, anti-codon arm, anti-codon loop, D-arm, D-loop, Ψ-arm, and Ψ-loop (Fig. S1). Notably, a few unique tRNAs such as *trnL* (Leucine), *trnS* (Serine) and *trnY* (Tyrosine) possessed an additional variable region that formed a stem with or without a loop (Fig. 2b–e). In addition, the *trnT* (UGU) possessed an internal loop in its acceptor arm (Fig. 2f) but similar internal loop was also detected in the *trnT* (UGU) of *Zingiber officinale* NC\_044775 and *Kaempferia elegans* NC\_040852 (Fig. S2).

All the genes were assigned into different gene groups that belonged to three major function categories including protein synthesis and self-replication, photosynthesis and other functions (Table 2). However, the distribution of these gene groups was uneven in the chloroplast genome. For instance, all genes in the photosystems I and II were found in the LSC region except for gene *psaC* that was located in the SSC region. While the rRNA genes were all found in the IR regions, the other tRNA genes and ribosomal protein genes were distributed throughout the chloroplast genome in all regions (Fig. 1).

**Amino acid and codon usage analyses** – There were a total of 27,743 codons in all the protein-coding genes that comprised of 61 types of codons encoding for 20 amino acids and three stop codons. Most of them possessed the canonical ATG start codon, except for genes *ndhD* with ATC as its start codon, and *rpl2* and *petB* with ATA as their start codon (Table S3). In addition, the amino acid leucine had the highest usage, while the amino acid tryptophan was the

**Table 1** Base composition of the *Boesenbergia rotunda* chloroplast genome

Region	Position	T (U) (%)	C (%)	A (%)	G (%)	GC content (%)	Length (bp)
Total		32.3	18.3	31.7	17.7	36.0	163,817
Inverted repeat A (IRA)		30.1	21.3	28.8	19.8	41.1	29,746
Inverted repeat B (IRB)		33.7	17.4	32.4	16.5	33.9	29,746
Large single copy (LSC)		28.8	19.8	30.1	21.3	41.1	88,302
Small single copy (SSC)		34.7	15.5	36.0	13.8	29.3	16,023
CDS		31.9	18.9	31.2	18.0	36.9	83,085
	First position	30.6	16.2	35.8	17.4	33.6	27,695
	Second position	31.0	18.3	31.6	19.1	37.4	27,695
	Third position	34.1	22.1	26.1	17.6	39.7	27,695



**Table 2** Gene contents in *Boesenbergia rotunda* chloroplast genome

Category	Gene groups	Gene names
Protein synthesis and self-replication	Transfer RNAs	<i>trnA-UGC*</i> (×2), <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnM-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC*</i> , <i>trnH-GUG</i> (×2), <i>trnI-CAU</i> (×2), <i>trnI-GAU*</i> (×2), <i>trnK-UUU*</i> , <i>trnL-CAA</i> (×2), <i>trnL-UAA*</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> (×2), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> (×2), <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> (×2), <i>trnV-UAC*</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
	Ribosomal RNAs	<i>rrn16</i> (×2), <i>rrn23</i> (×2), <i>rrn4.5</i> (×2), <i>rrn5</i> (×2)
	Ribosomal proteins (SSU)	<i>rps11</i> , <i>rps12**<sup>T</sup></i> (×2), <i>rps14</i> , <i>rps15</i> , <i>rps16*</i> , <i>rps18</i> , <i>rps19</i> (×2), <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i>
	Ribosomal proteins (LSU)	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2*</i> (×2), <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>
Photosynthesis	NADH-dehydrogenase	<i>ndhA*</i> , <i>ndhB*</i> (×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i> , <i>psbZ</i>
	Cytochrome b/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
	RubisCo Large subunit	<i>rbcL</i>
	Others	Translational initiation factor <i>infA</i>
Maturase	<i>matK</i>	
ATP-dependent protease	<i>clpP1**</i>	
Envelope membrane protein	<i>cemA</i>	
Subunit of Acetyl-CoA-carboxylase	<i>accD</i>	
c-Type cytochrome synthesis gene	<i>ccsA</i>	
Proteins of unknown function	<i>ycf1</i> (×2), <i>ycf2</i> (×2), <i>ycf3**</i> , <i>ycf4</i>	

\*Indicates genes containing one intron

\*\*Indicates genes containing two introns

<sup>T</sup>Indicates trans-spliced gene

×2Indicates genes with two copies

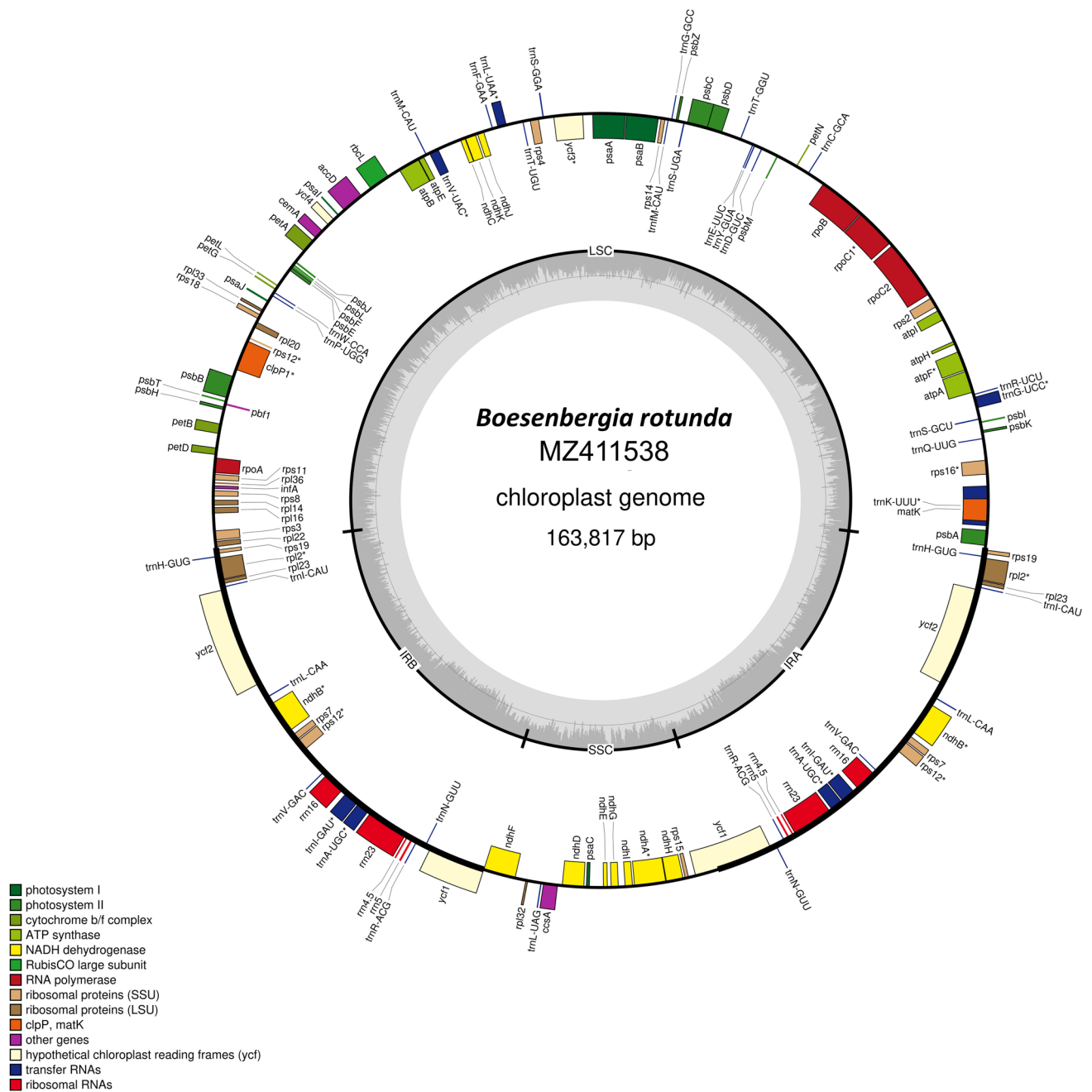
least used in the chloroplast genome (Fig. S3, Table S4). Additional relative synonymous codon usage (RSCU) was performed to assess the bias of synonymous codon usage in these protein genes by calculating the ratio of the observed frequency over the expected frequency for the codons (Sharp and Li 1986). The analysis revealed an uneven usage of synonymous codons in most amino acids, except for the codons of methionine and tryptophan that were encoded by AUG and UGG, respectively (RSCU = 1). Notably, the amino acid arginine showed a strong bias on codon AGA (RSCU > 2), slight bias for AGG and CGA (RSCU > 1), over the codons CGC, CGG and CGU (RSCU < 1) (Fig. 3). The analysis also revealed a biased usage of A/U than G/C at the third codon position in most amino acids of the protein-coding genes of *B. rotunda* chloroplast genome (Fig. 3).

**Analysis on SSRs and long repeats** – Our analysis detected a total of 108 SSRs in the *B. rotunda* chloroplast genome, and they were mostly borne on the LSC region (Table S5). The mononucleotide repeats are the most abundant SSRs in the chloroplast genome, and they mostly consisted of A/T mononucleotide repeats (Fig. 4). Similarly,

the 30 dinucleotide, nine trinucleotide, 16 tetranucleotide and three pentanucleotide SSRs detected were all AT-rich (Fig. 4a). These findings were consistent with the high overall AT content (74%) of the chloroplast genome.

Repeating DNA sequences with a length of 30 bp or longer are categorized as long repeats. Sixty-one long repeat structures were detected in the chloroplast genome of *B. rotunda* (Table S6). These included 25 palindromic, 24 forward and 12 reverse repeats (Fig. 4b). The majority of these long repeats were 30 to 39 bp in length (47 out of 61), while nine of them were 40 to 49 bp long, four were 50 to 59 bp long, and only one was 70 to 79 bp in length (Fig. 4b). While the majority of them were located in the LSC region, some long repeats were also detected on different regions (Table S6). There was no complementary repeat structure detected in the chloroplast genome.

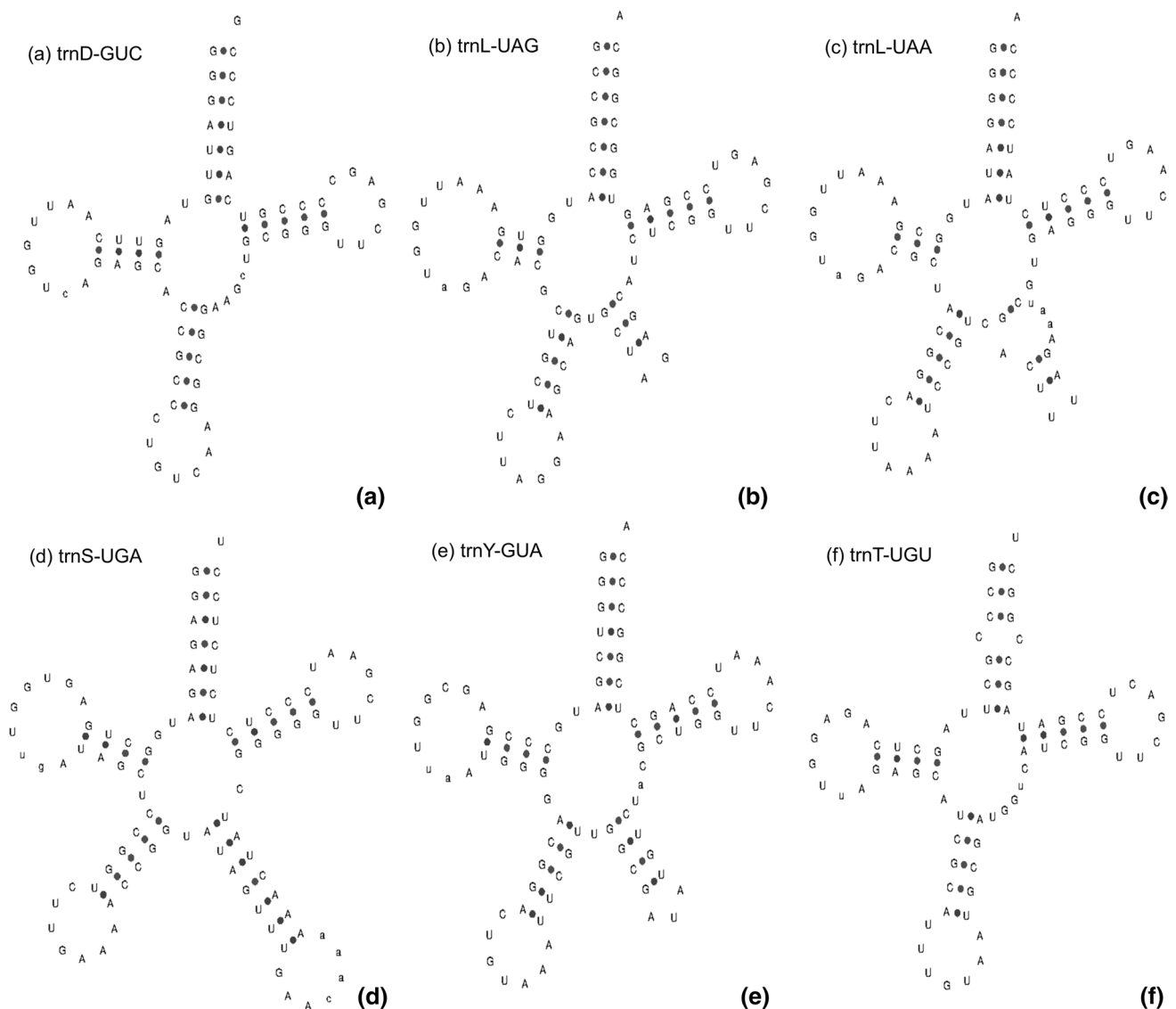
**Phylogenetic analysis** – The complete chloroplast genomes of *B. rotunda*, 49 ginger species in the family Zingiberaceae and an outgroup *Musa acuminata* HF677508 (Table S1) were included in a comprehensive phylogenetic analysis that involved the DNA sequences of 75 common



**Fig. 1** Gene map of the complete chloroplast genome of *Boesenbergia rotunda*. Genes are assigned according to color codes based on their category into gene groups. Genes drawn on the inside of the outer circle are transcribed counter-clockwise, while the genes drawn on the outside of the outer circle are transcribed clockwise. The dark gray region of the innermost circle corresponds to GC content, whereas the light gray region corresponds to AT content. *LSC* Large single-copy region, *SSC* small single-copy region, *IRA* inverted repeat region A, *IRB* inverted repeat region B

protein-coding genes (a total concatenated alignment length of 61,584 bp). The best-fitting substitution models for each gene partition were chosen using ModelFinder (Table S7). In both the ML and Bayesian inference (BI) phylogenetic trees, all the genera were delineated into distinct lineages except for the genus *Curcuma* (Fig. 5, Fig. S4). All the

*Curcuma* species formed a large cluster in the phylogenetic tree, except for *Curcuma flaviflora* S.Q.Tong that was placed in a clade with the *Zingiber* species (Fig. 5). A high ( $\geq 62$ ) to maximal bootstrap support on most nodes of the ML tree and a high ( $\geq 0.65$ ) to maximal Bayesian posterior probability on most nodes of the BI tree also indicated the confident



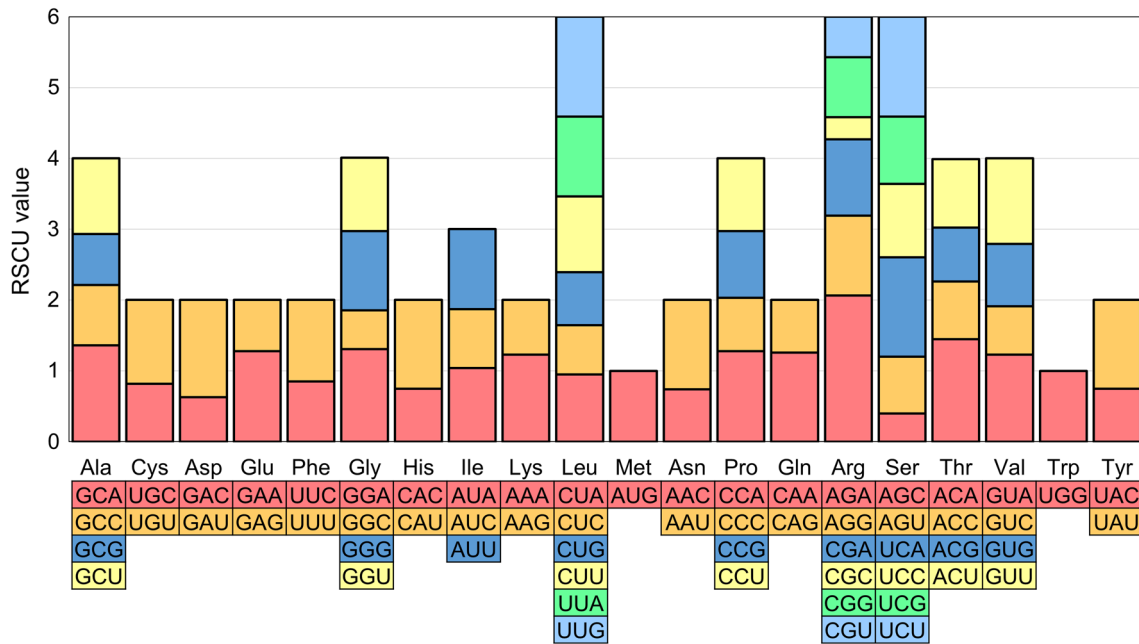
**Fig. 2** Structures of selected tRNAs from the chloroplast genome of *Boesenbergia rotunda*, **a** a canonical tRNA structure; **b** trnL-UAG; **c** trnL-UAA; **d** trnS-UGA; **e** trnY-GUA and **f** trnT-UGU

robustness of the analyses (Fig. 5, Fig. S4). All the taxa were also differentiated from the outgroup *Musa acuminata* HF677508.

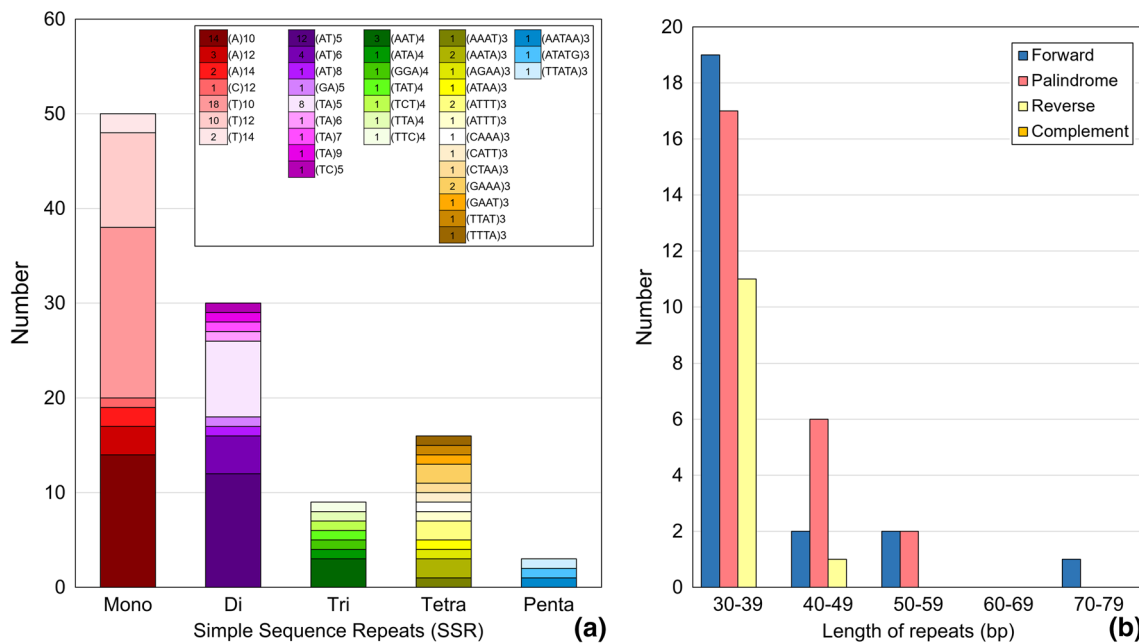
The tribe Zingiberaceae was represented by three subclades: (1) *Boesenbergia*, *Kaempferia* and *Zingiber* (and *C. flaviflora*); (2) *Hedychium*, *Roscoea*, *Pyrgophyllum* and *Cautleya*; and (3) *Curcuma* and *Stahlianthus* (Fig. 5, Fig. S4). *B. rotunda* and *B. kingii* Mood & L.M.Prince were in a clade basal to the genera *Kaempferia* and *Zingiber*. The tribes Zingiberaceae and Globbeae were placed in a large clade that constituted the subfamily Zingiberoideae. The Alpinoideae clade consisted of three subclades: (1) *Alpinia*; (2) *Amomum* and *Wurfbainia* and (3) *Lanxangia*. The complete chloroplast genome-based phylogenetic analysis

provided data to understand the evolutionary dynamics between the *Boesenbergia* species and various taxa of the family Zingiberaceae.

**Comparative analysis of chloroplast genome structure** – Twelve complete chloroplast genomes of closely related species in the family Zingiberaceae were included for a comparative analysis of chloroplast genome architecture using *B. rotunda* as the reference (Table S8). While the size of *B. rotunda* chloroplast genome was similar to the majority of the Zingiberaceae which is about 162 kbp to 164 kbp, the smallest and largest chloroplast genomes occurred in *R. humeana* Balf.f. & W.W.Sm. NC\_046582 (160,288 bp) and *Zingiber zerumbet* (L.) Roscoe ex Sm. NC\_049006 (169,183 bp), respectively (Table S8).



**Fig. 3** Relative synonymous codon usage (RSCU) of the protein-coding genes in the chloroplast genome of *Boesenbergia rotunda*. The y-axis indicated the value of RSCU

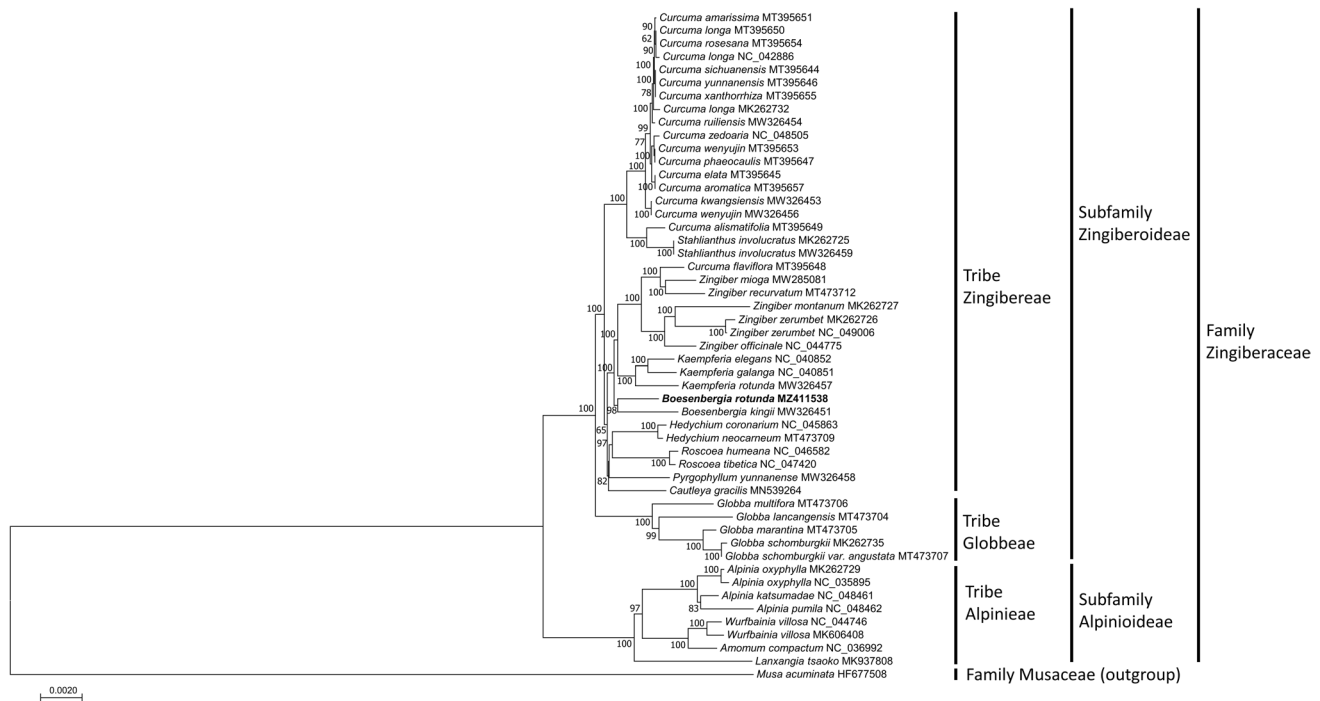


**Fig. 4** Distribution of sequence repeats in the chloroplast genome of *Boesenbergia rotunda*. **a** Simple sequence repeats (SSRs). The x-axis represents the types of SSRs and the y-axis indicates the number of SSR type. **b** Long repetitive sequence types including forward, palindrome, reverse and complement. The x-axis represents the length of these long repeats, and the y-axis indicates the number of long repeats

Interestingly, *Z. zerumbet* NC\_049006 and *Z. zerumbet* MK262726 possessed the smallest (86,709 bp) and the largest (89,161 bp) LSC regions (Table S8). It was also noteworthy that the smallest SSC region (7,515 bp) and

the largest IR region (74,960 bp) occurred in *Z. zerumbet* NC\_049006, while the largest SSC region (18,788 bp) and the smallest IR region (53,704 bp) occurred in *R. humeana* NC\_046582 (Table S8). The difference between the two





**Fig. 5** Phylogenetic relationships of *Boesenbergia rotunda* with the other species in the family Zingiberaceae. The *Musa acuminata* HF677508 was used as outgroup. The nucleotide substitution models for each gene partition (75 protein coding genes) were determined using ModelFinder. The phylogenetic tree was constructed using maximum likelihood method and the numbers at the nodes indicated bootstrap support values based on 10,000 replicates

*Z. zerumbet* chloroplast genomes corroborated that a large intraspecific variation may occur in terms of chloroplast genome region size in the family Zingiberaceae (Li et al. 2020; Qi et al. 2020).

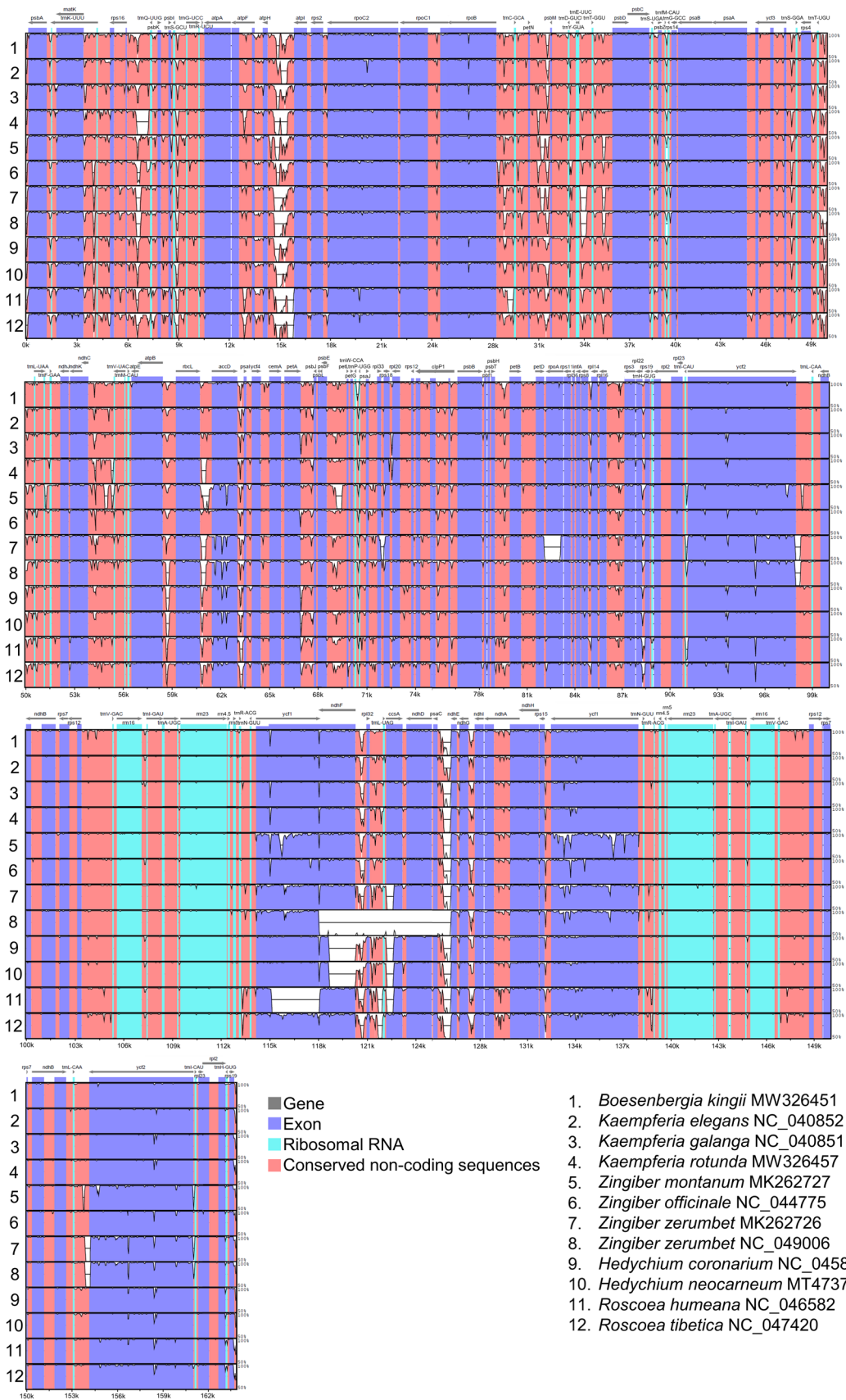
A comparison of sequence identity was carried out using mVISTA with the annotated chloroplast genome of *B. rotunda* as the reference, followed by an analysis on nucleotide variability ( $P_i$ ) to assess the variable regions in the chloroplast genomes. In general, the chloroplast genome alignment and visualization revealed little sequence variation in the coding region of protein exons and RNAs, while large variations were mostly observed in the non-coding regions. When compared to closely related species, the *B. rotunda* chloroplast genome had less sequence variation in both the IR regions (positions from about 88 kbp to 118 kbp and 134 kbp to 163 kbp) (Fig. 6). Several LSC regions, including 3–10 kbp (corresponding to non-coding regions between *matK* and *trnG-UCC*), 13–16 kbp (*atpF* to *atpI*), 28–36 kbp (*rpoB* to *psbD*), 49–51 kbp (*rps4* to *trnF-GAA*), and the non-coding regions (at 54–76 kbp) were highly variable as compared to the chloroplast genomes of closely related species (Fig. 6). These were consistent with the nucleotide variability analysis for which the lowest variability ( $P_i < 0.008$ ) occurred in both the IR regions, followed by the larger variability ( $P_i > 0.012$ ) in the LSC region (Fig. 7, Table S9). The same analysis also revealed several regions

with remarkably high nucleotide variability such as the *psbK-psbI*, *trnT-GGU-psbD*, *rbcL-accD*, *ndhF-rpl32* and *ycfI* gene regions (Fig. 7).

Notably, a very high sequence variation was also observed at the ending region of IRB and the starting region of SSC regions, at position about 120–128 kbp (*ndhF* to *ndhI*) (Fig. 6). These were also partly caused by sequence identity gaps in the SSC region of *Z. zerumbet* NC\_049006 and the intercepting IRB and SSC regions of *R. humeana* NC\_046582. Further investigation revealed duplication of genes including *rps15*, *ndhH*, *ndhA*, *ndhI*, *ndhG* and *ndhE*, in addition to inversion of these duplicated genes together with the genes *psaC*, *ndhD*, *ccsA*, *trnL-UAG* and *rpl32* at the SSC region of *Z. zerumbet* NC\_049006 (Fig. S5). In the case of *R. humeana* NC\_046582, the sequence identity gap was caused by a severely truncated *ycfI* gene that occurred at the end of IRB region (115–118 kbp) (Fig. S6). The variations in the SSC region were also indicated by the highest nucleotide variability ( $P_i > 0.04$ ) (Fig. 7, Table S9).

## 4 Discussion

In this study, the chloroplast genome of *B. rotunda* was obtained from a hybrid assembly of sequence data generated from both Illumina and Nanopore platforms. This approach



**Fig. 6** Sequence identity plot comparing the chloroplast genome of *Boesenbergia rotunda* and closely related species in the family Zingiberaceae. The vertical axis indicates the percentage of identity (50–100%), while the horizontal axis indicates the position in the chloroplast genome. The regions for gene, exon, RNA and non-coding sequence are color-coded, while the white region represents sequence variation among the species

overcome the challenges of Illumina short read sequencing data in correctly assembling especially the regions of inverted repeat of the chloroplast genome. The use of Nanopore long read sequencing generated sufficient information to completely cover the inverted repeat regions. Studies have shown that a hybrid approach enabled assembly of chloroplast genomes at higher accuracy than long- and short-reads only assemblies (Wang et al. 2018b; Guo et al. 2021).

Various analysis revealed very similar gene features and contents in the chloroplast genomes of *B. rotunda* and closely related Zingiberaceae taxa (Table S8). The localization of *matK* within the *trn-UUU* was not unique to *B. rotunda* as it had also been described in other ginger genera including *Zingiber* (Cui et al. 2019; Li et al. 2020) and *Kaempferia* (Li et al. 2019). Besides, our study detected a few tRNAs with an additional variable region that formed a stem with or without a loop in the *trnL* (Leucine), *trnS* (Serine) and *trnY* (Tyrosine) (Fig. 2b to e). These non-typical tRNA structures have also been reported in cotton and rice plant (Mohanta et al. 2019). The variable arm regions were thought to contribute to the tRNA structural variation and are crucial for the interaction with the D-arm and  $\Psi$ -arm as well as to maintain the tRNA structure (Wang et al. 2018a; Zhang et al. 2021).

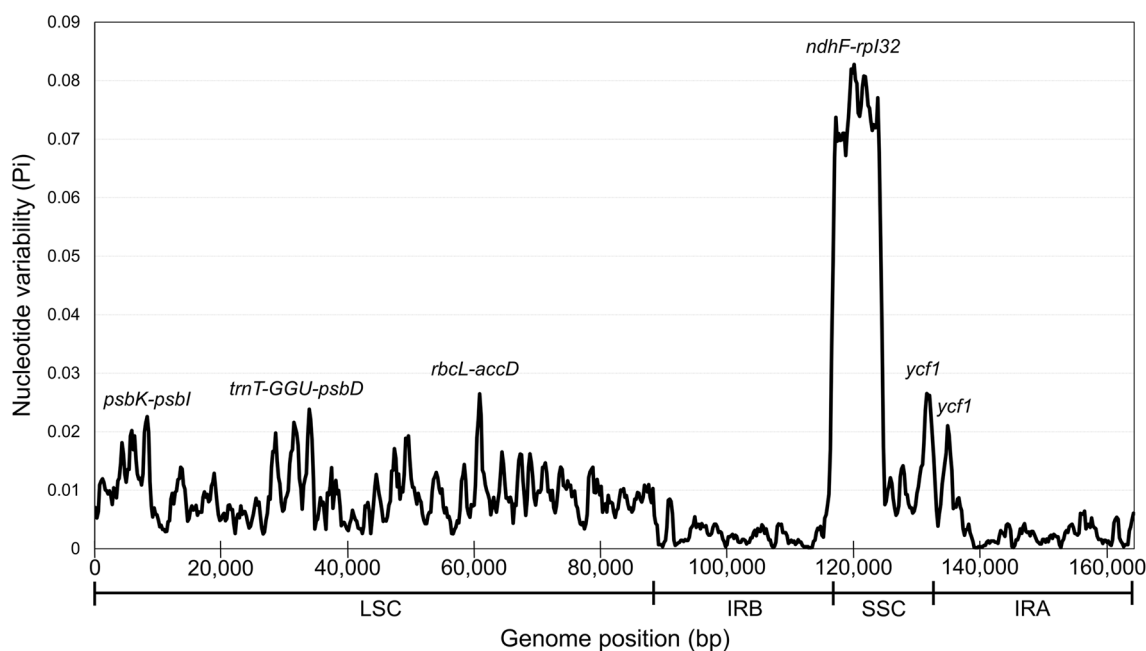
The *B. rotunda* chloroplast genome harbored genes with atypical start codons such as *ndhD* that had ATC as its start codon and *rpl2* and *petB* that possessed ATA as their start codon (Table S3). The use of ATC as the start codon of *ndhD* had been reported in the chloroplast genome of *Amomum compactum* Sol. ex Maton NC\_036992 (Wu et al. 2018), while the genes *rpl2* and *petB* had been reported with ATA as their start codon in the chloroplast genomes of *Alpinia oxyphylla* Miq. (Gao et al. 2019) and *Aquilaria sinensis* (Lour.) Gilg (Wang et al. 2016). A RSCU analysis revealed a biased codon usage in the genes of chloroplast genome of *B. rotunda* which was also reported in the genera *Zingiber* (Cui et al. 2019; Gao et al. 2019) and *Kaempferia* (Li et al. 2019). A biased usage of the synonymous codons in the chloroplast genome elevates the translational speed and reduces the costs of proofreading by using codons which match common tRNAs efficiently but the phenomenon might indicate a selection pressure (Xu et al. 2011; Duan et al. 2021). As the organelle and nuclear genes exhibit different features of codon usage, the results will be useful in future whole genome analysis to distinguish the gene contents

between the nuclear, mitochondrial and chloroplast genomes (Xu et al. 2011).

In addition, the chloroplast genome of *B. rotunda* harbored various SSRs and long repeats. The SSRs are tandem repeats that usually consist of one to six nucleotides per unit that are distributed throughout the chloroplast genome which are useful molecular markers to study genetic variation (Powell et al. 1995). Before the advent of next-generation sequencing technologies, it was difficult to develop SSRs-based molecular markers as it relied heavily on PCR analysis (Taheri et al. 2019). The findings on SSRs reported in this study are thus crucial for population genetic analysis of the family Zingiberaceae. Meanwhile, a total of sixty-one long repeat structures were detected in the chloroplast genome of *B. rotunda* (Table S6). This was slightly lesser than *B. kingii* (with 64 long repeats) (Liang and Chen 2021), but higher than many closely related *Zingiber* species (Cui et al. 2019; Li et al. 2020) and *Kaempferia* species (Li et al. 2019) (42–50 long repeats). The presence of these long repeats could influence the architecture of a plant chloroplast genome, causing duplication and rearrangement which could play an important role for evolutionary and phylogenetic analysis (Nie et al. 2012; Park et al. 2017). However, no gene duplication or genome rearrangement was detected in the chloroplast of these ginger taxa in this study.

The genus *Boesenbergia* was once regarded as a member in the same tribe Hedychieae that included the genus *Hedychium* based on morphological classification (Kress et al. 2002). The preliminary identification of *B. rotunda* and the Zingiberaceae species was still dependent on morphological characteristics (Saensouk et al. 2016). The availability of complete chloroplast genomes of *B. rotunda* and closely related Zingiberaceae taxa permitted genome-wide alignment and analysis on nucleotide variability to identify the polymorphic regions in the chloroplast genomes. These information provided insights into the suitability of individual chloroplast genes for phylogenetic inference and molecular identification of the Zingiberaceae species. More importantly, phylogenetic analysis based on all the common chloroplast protein-coding genes revealed a robust evolutionary relationship between *B. rotunda* and the other ginger species.

Before the complete chloroplast genomes were available, the molecular phylogenetic analysis of ginger species relied heavily on individual chloroplast genes such as *matK* (Kress et al. 2005; Saha et al. 2020) and *trnL* (UAA) 5' exon to *trnF* (GAA) (Ngamriabsakul et al. 2003). However, these single gene-based approaches did not provide a high resolution in differentiation of closely related species. In Kress et al. (2002), *Boesenbergia pulcherrima* (Wall.) Kuntze and *B. rotunda* formed two separate clades in a phylogenetic analysis based on *matK* and internal transcribed spacer (ITS) region. Although the chloroplast genome of *B. pulcherrima*



**Fig. 7** Nucleotide variability (Pi) of the whole chloroplast genome of *Boesenbergia rotunda* and closely related species in the family Zingiberaceae. Window length: 800 bp; step size: 200 bp; x-axis: Position of the midpoint of a window; y-axis: nucleotide diversity of each window

is not available, our analysis revealed a significant distance between *B. rotunda* and *B. kingii*, as well as a well-resolved phylogeny of the family Zingiberaceae which extended across multiple subfamilies and tribes (Figs. 5, S4). Nevertheless, a more extensive taxon sampling is necessary to reconstruct a robust and up-to-date phylogeny of the family.

In conclusion, this study reported the complete chloroplast genome of *B. rotunda* and compared it to other Zingiberaceae taxa. Various analyses described the characteristics of the chloroplast genome including the GC content, gene content, amino acid and codon usage, the presence of SSRs and long repeats. While a comparative analysis revealed a high sequence identity among the compared chloroplast genomes, several highly variable regions, gene duplication and truncation were identified among the ginger species. More importantly, a phylogenetic analysis based on all the protein-coding gene sequences was useful in resolving the relationships among the Zingiberaceae species. This study provided essential chloroplast genome information of *B. rotunda* that could be a notable reference for future species identification, phylogenetic and diversity research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40415-022-00845-w>.

**Acknowledgements** We thank Dr. Yong Kien Thai (University of Malaya, Malaysia) for providing assistance in voucher specimen deposit. This work was supported by grant from Mr. Yiang Kwai Sang (PV005-2017), the University of Malaya Research Grants (FRGS grant FP022-2018A), University of Malaya High Impact Research Grants (UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/CHAN/14/1, Grant

No. H-50001-A000027; UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/CHAN/01, Grant No. A-000001-50001) awarded to KGC.

**Author contributions** KGC conceived the study, reviewed and edited the draft of the paper. YJML performed the experiments, data analysis and drafted the manuscript. KOC performed the data analysis and drafted the manuscript. HSY and SLS validated the analysis, reviewed and edited the manuscript draft. All authors read and approved the final manuscript.

**Funding** Not applicable.

**Availability of data and materials** The complete chloroplast genome sequence of *B. rotunda* was deposited in the GenBank under the Accession Number MZ411538.

## Declarations

**Conflict of interest** The authors declare that there are no conflicts of interest.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

## References

- Antipov D, Korobeynikov A, McLean JS, Pevzner PA (2016) hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32:1009–1015

- Atun S, Handayani S, Rakhmawati A (2018) Potential bioactive compounds isolated from *Boesenbergia rotunda* as antioxidant and antimicrobial agents. *Pharmacogn J* 10:513–518
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Branney TME (2005) Hardy ginger: including *Hedygium*, *Roscoeia*, and *Zingiber*. Timber Press, Kuala Lumpur
- Break MKB, Chiang M, Wiart C, Chin C-F, Khoo ASB, Khoo T-J (2021) Cytotoxic activity of *Boesenbergia rotunda* extracts against nasopharyngeal carcinoma cells (HK1). cardamonin, a *Boesenbergia rotunda* constituent, inhibits growth and migration of HK1 cells by inducing caspase-dependent apoptosis and G2/M-phase arrest. *Nutr Cancer* 73:473–483
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973
- Chan PP, Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M (ed) Gene prediction methods in molecular biology, vol 1962. Humana, New York, pp 1–14
- Chen J, Xia N-H (2019) A taxonomic revision of Chinese *Boesenbergia* (Zingiberaceae), with a new record. *Phytotaxa* 424:217–231
- Cui Y, Nie L, Sun W, Xu Z, Wang Y, Yu J, Song J, Yao H (2019) Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the family Zingiberaceae based on the complete chloroplast genome. *Plants* 8:283
- Daniell H, Lin C-S, Yu M, Chang W-J (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:1–29
- Duan H, Zhang Q, Wang C, Li F, Tian F, Lu Y, Hu Y, Yang H, Cui G (2021) Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ* 9:e10787
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucl Acids Res* 32:W273–W279
- 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:e0218817
- Green BR (2011) Chloroplast genomes of photosynthetic eukaryotes. *Plant J* 66:34–44
- Greiner S, Lehwerk P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucl Acids Res* 47:W59–W64
- Guo Y-Y, Yang J-X, Li H-K, Zhao H-S (2021) Chloroplast genomes of two species of *Cypripedium*: expanded genome size and proliferation of AT-biased repeat sequences. *Front Plant Sci* 12:609729
- Hooker JD (1875) The flora of British India. L. Reeve
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
- Inglis PW, Pappas MdCR, Resende LV, Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS ONE* 13:e0206085
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589
- Kanjanasirirat P, Suksatu A, Manopwisedjaroen S, Munyoo B, Tuchinda P, Jearawuttanakul K, Seemakhan S, Charoensuththivarakul S, Wongtrakoongate P, Rangkasene N (2020) High-content screening of Thai medicinal plants reveals *Boesenbergia rotunda* extract and its component Panduratin A as anti-SARS-CoV-2 agents. *Sci Rep* 10:1–12
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- Kono N, Arakawa K (2019) Nanopore sequencing: review of potential applications in functional genomics. *Dev Growth Differ* 61:316–326
- Kress WJ, Prince LM, Williams KJ (2002) The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Am J Bot* 89:1682–1696
- Kress WJ, Liu AZ, Newman M, Li QJ (2005) The molecular phylogeny of *Alpinia* (Zingiberaceae): a complex and polyphyletic genus of gingers. *Am J Bot* 92:167–178
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucl Acids Res* 29:4633–4642
- Larsen K (1996) A preliminary checklist of the Zingiberaceae of Thailand. *Thai for Bull (botany)* 24:35–49
- Larsen K, Ibrahim H, Khaw S, Saw L (1999) Gingers of peninsular Malaysia and Singapore. Natural History Publications (Borneo)
- Li D-M, Zhao C-Y, Liu X-F (2019) Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: molecular structures and comparative analysis. *Molecules* 24:474
- Li D-M, Ye Y-J, Xu Y-C, Liu J-M, Zhu G-F (2020) Complete chloroplast genomes of *Zingiber montanum* and *Zingiber zerumbet*: genome structure, comparative and phylogenetic analyses. *PLoS ONE* 15:e0236590
- Liang H, Chen J (2021) Comparison and phylogenetic analyses of nine complete chloroplast genomes of Zingiberaceae. *Forests* 12:710
- Liew YJM, Lee YK, Khalid N, Abd Rahman N, Tan BC (2020) Enhancing flavonoid production by promiscuous activity of prenyltransferase, BrPT2 from *Boesenbergia rotunda*. *PeerJ* 8:e9094
- Liew YJM, Lee YK, Khalid N, Abd Rahman N, Tan BC (2021) Cell-free expression of a plant membrane protein BrPT2 from *Boesenbergia rotunda*. *Mol Biotechnol* 63:316–326
- Mohan S, Hobani YH, Shaheen E, Abou-Elhamd AS, Alhazmi HA, Abdelwahab SI (2020) Ameliorative effect of Boesenbergin A, a chalcone isolated from *Boesenbergia rotunda* (Fingerroot) on oxidative stress and inflammation in ethanol-induced gastric ulcer in vivo. *J Ethnopharmacol* 261:113104
- Mohanta TK, Khan AL, Hashem A, Allah EFA, Yadav D, Al-Harrasi A (2019) Genomic and evolutionary aspects of chloroplast tRNA in monocot plants. *BMC Plant Biol* 19:1–24
- Ngamriabsakul C, Newman MF, Cronk QCB (2003) The phylogeny of tribe Zingiberaceae (Zingiberaceae) based on ITS (nrDNA) and trnL-F (cpDNA) sequences. *Ethin J Bot* 60:483–507
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* 7:e36869
- Park I, Yang S, Choi G, Kim WJ, Moon BC (2017) The complete chloroplast genome sequences of *Aconitum pseudolaeve* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the *Aconitum* Subgenus *Lycocotnum*. *Molecules* 22:2012
- Powell W, Morgante M, McDevitt R, Vendramin G, Rafalski J (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl Acad Sci* 92:7759–7763



- Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes de novo assembler. *Curr Protoc Bioinform* 70:e102
- Qi H, Yun Y, Chen X, Ren J, Sun X, Zheng D (2020) Characterization of the complete chloroplast genome of *Zingiber zerumbet*. *Mitochondr DNA Part B* 5:2629–2630
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302
- Saensouk S, Saensouk P, Pasorn P, Chantaranonthai P (2016) Diversity and uses of Zingiberaceae in Nam Nao National Park, Chaiyaphum and Phetchabun provinces, Thailand, with a new record for Thailand. *Agric Nat Resour* 50:445–453
- Saha K, Sinha RK, Sinha S (2020) Distribution, cytology, genetic diversity and molecular phylogeny of selected species of Zingiberaceae—a review. *Feddes Repertorium* 131:58–68
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Sharp PM, Li W-H (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* 24:28–38
- Stortchevoi A, Kamelamela N, Levine SS (2020) SPRI Beads-based size selection in the range of 2–10kb. *J Biomol Tech* 31:7
- Sudianto E, Wu C-S, Leonhard L, Martin WF, Chaw S-M (2019) Enlarged and highly repetitive plastome of *Lagarostrobos* and plastid phylogenomics of Podocarpaceae. *Mol Phylogenet Evol* 133:24–32
- Sullivan MJ, Zakour NLB, Forde BM, Stanton-Cook M, Beatson SA (2015) Contiguity: contig adjacency graph construction and visualisation. *PeerJ PrePrints* 3:e1037v1
- Sun Y, Shang L, Zhu Q-H, Fan L, Guo L (2021) Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci* 27:391–401
- Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SPO, Teo CH, Sahebi M, Azizi P (2019) De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. *Sci Rep* 9:1–14
- Tan EH, Lee Y-K, Chee C-F, Heh C-H, Wong S-M, Thio L-PC, Foo G-T, Khalid N, Noorsaadah AR, Saiful Anuar K (2012) *Boesenbergia rotunda*: from ethnomedicine to drug discovery. *Evid Based Complement Altern Med* 2012:473637
- Tanabe AS (2007) Kakusan: a computer program to automate the selection of a nucleotide substitution model and the configuration of a mixed model on multilocus data. *Mol Ecol Notes* 7:962–964
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq—versatile and accurate annotation of organelle genomes. *Nucl Acids Res* 45:W6–W11
- Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP (2018) MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Res* 28:266–274
- Wang C, Zhang Y, Ding H, Song M, Yin J, Yu H, Li Z, Han L, Zhang Z (2021) Authentication of *Zingiber* species based on analysis of metabolite profiles. *Front Plant Sci* 12:705446
- Wang W, Chen X, Wolin SL, Xiong Y (2018a) Structural basis for tRNA mimicry by a bacterial Y RNA. *Structure* 26(1635–1644):e1633
- Wang W, Schalamun M, Morales-Suarez A, Kainer D, Schwessinger B, Lanfear R (2018b) Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genom* 19:1–15
- Wang Y, Zhan D-F, Jia X, Mei W-L, Dai H-F, Chen X-T, Peng S-Q (2016) Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. *Front Plant Sci* 7:280
- Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297
- Wu M-l, Li Q, Xu J, Li X-w (2018) Complete chloroplast genome of the medicinal plant *Amomum compactum*: gene organization, comparative analysis, and phylogenetic relationships within Zingiberales. *Chin Med* 13:1–12
- Xu C, Cai X, Chen Q, Zhou H, Cai Y, Ben A (2011) Factors affecting synonymous codon usage bias in chloroplast genome of *oncidium gower ramsey*. *Evol Bioinform* 7:EBO.S8092
- Zhang T-T, Yang Y, Song X-Y, Gao X-Y, Zhang X-L, Zhao J-J, Zhou K-H, Zhao C-B, Li W, Yang D-G (2021) Novel structural variation and evolutionary characteristics of chloroplast tRNA in *Gossypium* plants. *Genes* 12:822

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.