


RESEARCH ARTICLE

Open Access

# The chloroplast genomes of four *Bupleurum* (Apiaceae) species endemic to Southwestern China, a diversity center of the genus, as well as their evolutionary implications and phylogenetic inferences



Rong Huang<sup>1†</sup>, Xuena Xie<sup>1†</sup>, Aimin Chen<sup>1</sup>, Fang Li<sup>1</sup>, Enwei Tian<sup>1</sup> and Zhi Chao<sup>1,2,3\*</sup> 

## Abstract

**Background:** As one of the largest genera in Apiaceae, *Bupleurum* L. is well known for its high medicinal value. The genus has frequently attracted the attention of evolutionary biologist and taxonomist for its distinctive characteristics in the Apiaceae family. Although some chloroplast genomes data have been now available, the changes in the structure of chloroplast genomes and selective pressure in the genus have not been fully understood. In addition, few of the species are endemic to Southwest China, a distribution and diversity center of Chinese *Bupleurum*. Endemic species are key components of biodiversity and ecosystems, and investigation of the chloroplast genomes features of endemic species in *Bupleurum* will be helpful to develop a better understanding of evolutionary process and phylogeny of the genus. In this study, we analyzed the sequences of whole chloroplast genomes of 4 Southwest China endemic *Bupleurum* species in comparison with the published data of 17 *Bupleurum* species to determine the evolutionary characteristics of the genus and the phylogenetic relationships of Asian *Bupleurum*.

**Results:** The complete chloroplast genome sequences of the 4 endemic *Bupleurum* species are 155,025 bp to 155,323 bp in length including a SSC and a LSC region separated by a pair of IRs. Comparative analysis revealed an identical chloroplast gene content across the 21 *Bupleurum* species, including a total of 114 unique genes (30 tRNA genes, 4 rRNA genes and 80 protein-coding genes). Chloroplast genomes of the 21 *Bupleurum* species showed no rearrangements and a high sequence identity (96.4–99.2%). They also shared a similar tendency of SDRs and SSRs, but differed in number (59–83). In spite of their high conservation, they contained some mutational hotspots, which can be potentially exploited as high-resolution DNA barcodes for species discrimination. Selective pressure analysis showed that four genes were under positive selection. Phylogenetic analysis revealed that the 21 *Bupleurum* formed two major clades, which are likely to correspond to their geographical distribution.

\* Correspondence: [chaozhi@smu.edu.cn](mailto:chaozhi@smu.edu.cn)

<sup>†</sup>Rong Huang and Xuena Xie contributed equally to this work.

<sup>1</sup>Department of Pharmacy, Zhujiang Hospital, Southern Medical University, Guangzhou 510282, China

<sup>2</sup>Faculty of Medicinal Plants and Pharmacognosy, School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, China

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



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

**Conclusions:** The chloroplast genome data of the four endemic *Bupleurum* species provide important insights into the characteristics and evolution of chloroplast genomes of this genus, and the phylogeny of *Bupleurum*.

**Keywords:** *Bupleurum*, Chloroplast genome, Comparative analysis, Phylogenetic analyses

## Background

*Bupleurum* L., with more than 180 species, represents one of the largest genera of the family Apiaceae and is distributed in the north temperate zone (mainly in Eurasia, the Mediterranean, North Africa, Asia and North America) [1, 2]. Different from other genera of Apiaceae, life forms in *Bupleurum* vary greatly, ranging from herbs to shrubs [3]. The genus has frequently attracted the attention of evolutionary biologists and taxonomists for its distinctive characteristics in the Apiaceae family [1–6].

*Bupleurum* is easily distinguished from other genera of the family for its simple and entire leaves as well as conspicuous bracts and bracteoles, which are almost unique morphology characteristics in the family [1, 4]. Molecular phylogenetic studies in the recent two decades based on plastid and nuclear markers suggest that it should be considered as an identifiable tribe [7–12]. Interspecific phylogeny of *Bupleurum* presents a long standing problem in the systematics. The genus shows broad intra-specific morphological variations with poorly defined inter-specific boundaries, making the taxonomy based on traditional classification systems extremely difficult as increasing species being discovered [5, 6, 13]. In spite of the efforts at phylogenetic analysis in previous studies, some essential problems concerning the phylogenetic relationship of *Bupleurum* based on nrDNA (ITS) and various plastid sequences (e. g. *rps16*, *trnH-psbA*, and *matK*) still remain to be solved [1, 4, 13, 14].

Compared with nuclear and mitochondrion genome, gene density of chloroplast genome is larger and the evolution rate is moderate, and the segments with different evolution rates may serve for different research purposes [15–19]. The evolution of chloroplast genomes has long fascinated and puzzled evolutionary biologists. The understanding of the evolutionary connection among the plant species, the features they shared, and their differences from other taxonomic groups [20] all benefit from comparative analysis of whole chloroplast genomes that provide insights into the phylogenetic relationships and species evolution in different taxa [21–28]. In general, chloroplast genome has long been considered to be conserved and affected little by adaptive evolution in many genera [18–20]. However, rearrangements, differences in structure, size, gene content and order, and positively selected genes have been documented in some genera, such as *Amphilophium* [29], *Amorphophallus* [30] and the apiaceous genus *Ligusticum* [31]. For such a

diversified and wide-distributed genus like *Bupleurum*, we can not assert that there is absolutely no variation in chloroplast genome in a certain group without detailed study. It is important to investigate chloroplast genomes of the taxonomically significant genera, for understanding how infrageneric species are linked, what features are shared among them, and how they are different from other taxonomic groups [20, 29]. The advancement of the high-throughput sequencing technologies has drastically lowered the cost of analysis of the whole chloroplast genome sequences. Previous studies have reported the sequence data of chloroplast genomes for some species of *Bupleurum* [32–40], but unfortunately the changes in the structure of chloroplast genomes or the selective pressure in the genus were seldom addressed. As a result, the evolution of chloroplast of the genus is poorly understood. In the study by Li et al. [37], the chloroplast genomes of only five *Bupleurum* species were reported, but few of the species are endemic to Southwest China. Southwest China, which harbors an extremely high species diversity [41, 42] and is a distribution and diversity center of Chinese *Bupleurum* (ca. 21 species), including 12 species and 8 varieties endemic to China, 11 species and 5 varieties endemic to Southwest China. As key components of biodiversity and ecosystems, endemic species has long attracted the interest of ecologists and evolutionary biologists [43–47]. Investigation of the chloroplast genome features of endemic *Bupleurum* species may provide important insights into the evolution and phylogeny of the genus, especially the endemic *Bupleurum* species, thus helping to clarify the phylogenetic relationships and evolutionary aspects in the genus.

*Bupleurum shanianum*, *B. yunnanense*, *B. kweichowense* and *B. rockii* are endemic to Southwest China [2, 6]. *B. shanianum*, *B. yunnanense* and *B. kweichowense* are 3 perennial herbs with height ca. 6–35(–58) cm, 12–35 cm, 20–40 cm, respectively, while *B. rockii* has a relatively higher height ca. 60–100 cm. *B. shanianum* and *B. yunnanense* grow among grassy places, bushes, or under forests at altitudes of 3200–4400 m and 2500–5000 m, respectively, distributed in the alpine area of eastern Himalayas (Southeast Tibet), Sichuan and Yunnan Provinces of China. *B. rockii* grows in open forests and grasses on mountain slopes at altitudes of 1900–4200 m, occurring in Sichuan and northwest Yunnan Provinces of China. The documentation of *B. kweichowense* is

extremely poor and only a few collections from north-east Guizhou Province (Fanjing Shan) of China are available. It grows on gravelly slopes in sunny places at altitudes ca. 2100 m.

In this study, we used high-throughput sequencing technologies to sequence the chloroplast genomes of four *Bupleurum* species (*B. shanianum*, *B. yunnanense*, *B. kweichowense* and *B. rockii*) and assembled their whole chloroplast genomes. Previously published chloroplast genomes of 17 *Bupleurum* species [32–40], including herbs and one shrub (*B. dracaenoides*), were downloaded for comparative analysis, and the 21 species are distributed roughly in 3 regions (Southwest China; Northwest China; Northeast China, Korea and Japan; Additional file 1: Fig. S1 and Additional file 2: Table S1), which vary greatly in geomorphology and climate. Environmental heterogeneity plays an important role in evolutionary trajectories and ecological adaptation of species. It has been reported that the positive selection on some plastid genes (e. g. *clpP*, *ndhF* and *matK*) were observed on some plastid genes in some genera, which indicate that these genes might be subject to adaptive evolution in response to distinct ecological selective pressures [29–31]. However, to date, adaptive evolution of chloroplast genes in *Bupleurum* has not been fully understood. The sequence data of the chloroplast genomes provide a clue to the evolution and phylogeny of *Bupleurum*. Here, we attempted to answer the following questions: (1) Are there differences in the gene and structure of chloroplast genomes among *Bupleurum* species with different life forms and distributions? (2) Do the genes of *Bupleurum* species suffer positive selection under different habitats? We also constructed a phylogeny using 80 protein-coding genes (PCGs) of 21 *Bupleurum* species and 2 outgroups to explore the interspecific phylogenetic relationship of *Bupleurum* species in these regions.

## Results

### Chloroplast genome features of four *Bupleurum* species endemic to Southwest China

The sequences of the 4 *Bupleurum* chloroplast genomes ranged from 154,925 bp (*B. kweichowense*) to 155,323 bp (*B. yunnanense*), all having the typical quadripartite structure, comprising a SSC (17,478–17,575 bp) and a LSC (84,920–85,228 bp) region separated by a pair of IRs (52,572–52,649 bp) (Table 1, Fig. 1). The LSC regions exhibited the greatest standard deviation in sequence length (cv = 0.3%), followed by the IR regions (cv = 0.1%) and SSC regions (cv = 0.08%). The overall GC content was highly similar across the 4 chloroplast sequences (37.7–37.8%) (Table 1).

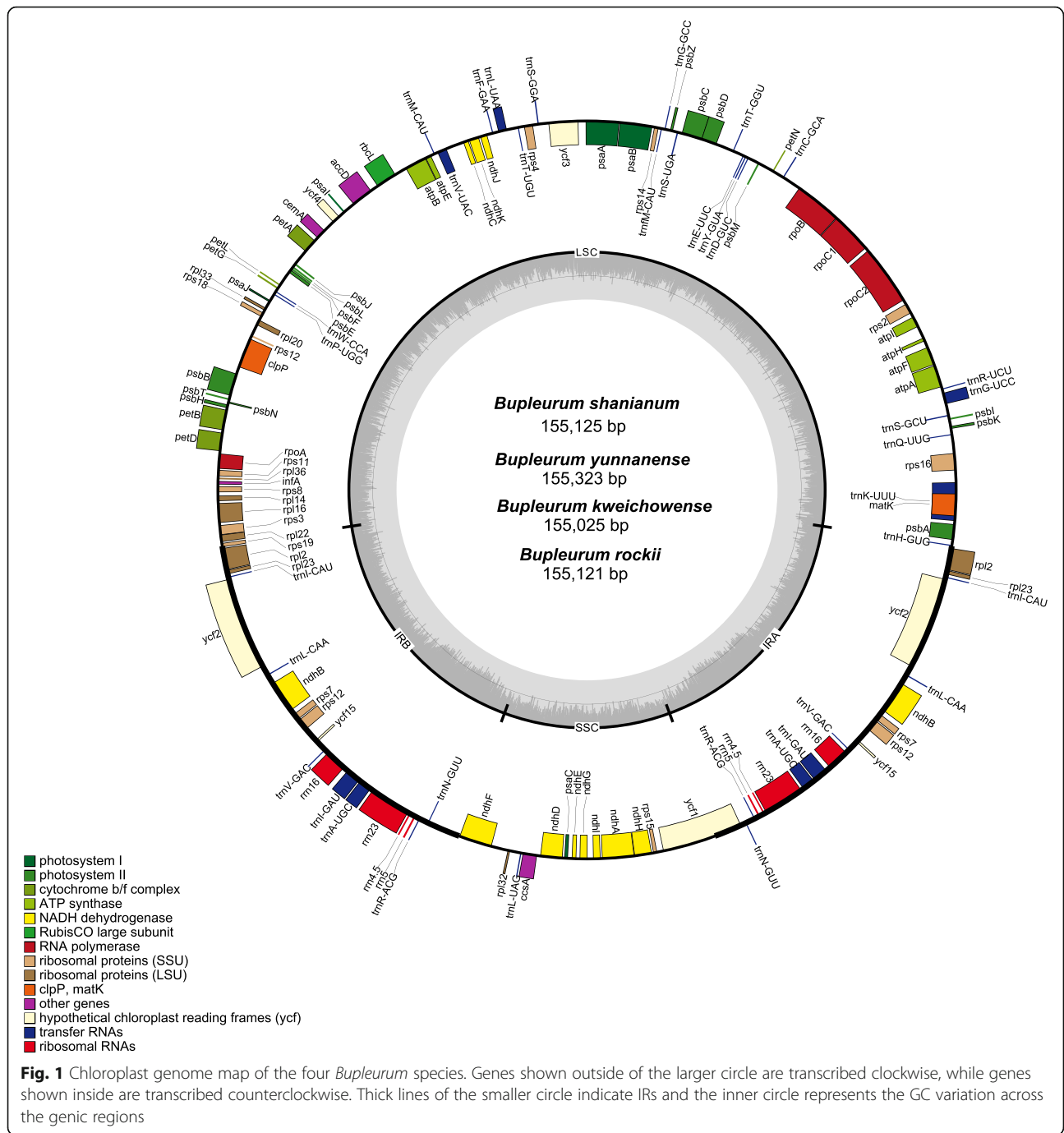
The chloroplast gene contents of the 4 *Bupleurum* species were identical (Table 1). Each *Bupleurum* chloroplast genome encoded a total of 114 unique genes, consisting of 30 tRNA genes, 4 rRNA genes and 80 protein-coding genes (PCGs) with the same gene order (Table 1). The SSC region contained 11 PCGs (*ndhF*, *rpl32*, *ccsA*, *ndhD*, *psaC*, *ndhE*, *ndhG*, *ndhI*, *ndhA*, *ndhH*, and *rps15*) and 1 tRNA (*trnL-UAG*), while the LSC region contained 60 PCGs and 22 tRNAs (Table 2). A total of 20 genes were duplicated in the IR regions, including 8 tRNAs (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnI-CAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), 8 PCGs (*rps7*, *rpl2*, *rpl23*, *ndhB*, *rps7*, *rps12*, *ycf2* and *ycf15*), and 4 rRNAs (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*) (Table 2). Eighteen genes contain introns, 15 of which contain a single intron, whereas the other 3 (*clpP*, *ycf3* and *rps12*) harbored 2 introns (Table 2). In this study, the incomplete copy of *ycf1* and *rps19* in the IR regions were regarded as pseudogenes.

### Comparative analysis of chloroplast genome structure of *Bupleurum*

The Mauve alignment analysis revealed that there was no rearrangement in coding and non-coding regions of the 21 *Bupleurum* chloroplast genomes

**Table 1** Summary of chloroplast features in four *Bupleurum* species

	<i>B. shanianum</i>	<i>B. yunnanense</i>	<i>B. kweichowense</i>	<i>B. rockii</i>
Total length (bp)	154,925–155,258	155,211–155,323	155,025	155,120–15,121
Total GC content (%)	37.8	37.8	37.7	37.8
IRa length (bp)	26,261–26,271	26,212–26,279	26,299	26,322
IRb length (bp)	26,261–26,300	26,212–26,279	26,299	26,327
SSC length (bp)	17,478	17,487–17,530	17,575	17,515
LSC length (bp)	84,921–85,208	85,148–85,228	84,920	84,959–84,960
Number unique genes	114	114	114	114
Protein coding	80	80	80	80
tRNA genes	30	30	30	30
rRNA genes	4	4	4	4



(Fig. 2), indicating that the chloroplast genomes were conserved. Among the 21 *Bupleurum* species, the genes *rpl22* and *rpl2* flanked the LSC/IRb junction, and gene *rps19* traversed the LSC and the IRb region (JLB line), with 50–82 bp occurring in the IR region (Fig. 3). On the other side of the IRb/SSC was the gene *ndhF*, which was 15–39 bp away from the IRb/SSC junction. The *ycf1* gene traversed the SSC and IRa region, with 1797–2140 bp occurring in the IR region (Fig. 3).

Simple sequence repeats (SSRs) analysis showed that total number of SSR loci ranged from 59 (*B. dracaenoides*) to 83 (*B. thianschanicum*). The patterns of SSRs distribution were similar among the 21 *Bupleurum*, as shown in Additional file 1: Fig. S2. Mono-nucleotides were the most frequent in the SSRs (61.5–71.2%), followed by di-nucleotides (10.1–19.7%). Tri-nucleotides and tetra-nucleotides were more frequent than penta- and hexa-nucleotides (Additional file 1: Fig. S2). Short dispersed repeats (SDRs) analysis showed that total

**Table 2** List of genes encoded in four *Bupleurum* plastomes

Gene Category	Genes	Number
Ribosomal RNAs	<i>rrn4.5</i> <sup>(×2)</sup> ; <i>rrn5</i> <sup>(×2)</sup> ; <i>rrn16</i> <sup>(×2)</sup> ; <i>rrn23</i> <sup>(×2)</sup>	4
Transfer RNAs	<i>trnA</i> -UGC <sup>(×2)a</sup> ; <i>trnC</i> -GCA; <i>trnD</i> -GUC; <i>trnE</i> -UUC; <i>trnF</i> -GAA; <i>trnI</i> M-CAU; <i>trnG</i> -GCC; <i>trnG</i> -UCC <sup>a</sup> ; <i>trnH</i> -GUG; <i>trnI</i> -CAU <sup>(×2)</sup> ; <i>trnI</i> -GAU <sup>(×2)</sup> <sup>a</sup> ; <i>trnK</i> -UUU <sup>a</sup> ; <i>trnL</i> -CAA <sup>(×2)</sup> ; <i>trnL</i> -UAA <sup>a</sup> ; <i>trnL</i> -UAG; <i>trnM</i> -CAU; <i>trnN</i> -GUU <sup>(×2)</sup> ; <i>trnP</i> -UGG; <i>trnQ</i> -UUG; <i>trnR</i> -ACG <sup>(×2)</sup> ; <i>trnR</i> -UCU; <i>trnS</i> -GCU; <i>trnS</i> -GGA; <i>trnS</i> -UGA; <i>trnT</i> -GGU; <i>trnT</i> -UGU; <i>trnV</i> -GAC <sup>(×2)</sup> ; <i>trnV</i> -UAC <sup>a</sup> ; <i>trnW</i> -CCA; <i>trnY</i> -GUA	37
Subunits of photosystem I	<i>psaA</i> ; <i>psaB</i> ; <i>psaC</i> ; <i>psal</i> ; <i>psaI</i> ; <i>ycf3</i> <sup>b</sup> ; <i>ycf4</i>	7
Subunits of 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>psbN</i> ; <i>psbT</i> ; <i>psbZ</i> ;	15
ATP-dependent protease subunit P	<i>clpP</i> <sup>b</sup>	1
Large subunit of rubisco	<i>rbcL</i>	1
NADH dehydrogenase	<i>ndhA</i> ; <i>ndhB</i> <sup>(×2)</sup> <sup>a</sup> ; <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>	12
Ribosomal protein (large subunit)	<i>rpl2</i> <sup>(×2)</sup> <sup>a</sup> ; <i>rpl14</i> ; <i>rpl16</i> <sup>a</sup> ; <i>rpl20</i> ; <i>rpl22</i> ; <i>rpl23</i> <sup>(×2)</sup> ; <i>rpl33</i> ; <i>rpl32</i> ; <i>rpl36</i>	11
Small subunit of ribosomal proteins	<i>rps2</i> ; <i>rps3</i> ; <i>rps4</i> ; <i>rps7</i> <sup>(×2)</sup> ; <i>rps8</i> ; <i>rps11</i> ; <i>rps12</i> <sup>(×2)b</sup> ; <i>rps14</i> ; <i>rps15</i> ; <i>rps16</i> <sup>a</sup> ; <i>rps18</i> ; <i>rps19</i>	14
DNA-dependent RNA polymerase	<i>rpoA</i> ; <i>rpoB</i> ; <i>rpoC1</i> ; <i>rpoC2</i>	4
Subunits of ATP synthase	<i>atpA</i> ; <i>atpB</i> ; <i>atpE</i> ; <i>atpF</i> <sup>a</sup> ; <i>atpH</i> ; <i>atpI</i>	6
C-type cytochrome synthesis gene	<i>ccsA</i>	1
Subunits of cytochrome b/f complex	<i>petN</i> ; <i>petA</i> ; <i>petL</i> ; <i>petG</i> ; <i>petB</i> <sup>a</sup> ; <i>petD</i> <sup>a</sup>	6
Envelop membrane protein	<i>cemA</i>	1
Maturase	<i>matK</i>	1
Hypothetical chloroplast reading frames	<i>ycf1</i> ; <i>ycf2</i> <sup>(×2)</sup>	5
Subunits of Acetyl-CoA-carboxylase	<i>accD</i>	1
Pseudogenes	<i>infA</i> ; <i>rps19</i> <sup>c</sup> ; <i>ycf1</i> <sup>c</sup> ; <i>ycf15</i> <sup>(×2)</sup>	4
Total	114 single-copy genes, 132 in total.	

(×2): Two gene copies in the IRs; <sup>a</sup> Gene containing one intron; <sup>b</sup> Gene containing two introns; <sup>c</sup> means the incomplete copy located in the IR of the gene straddling the IR and LSC/SSC regions

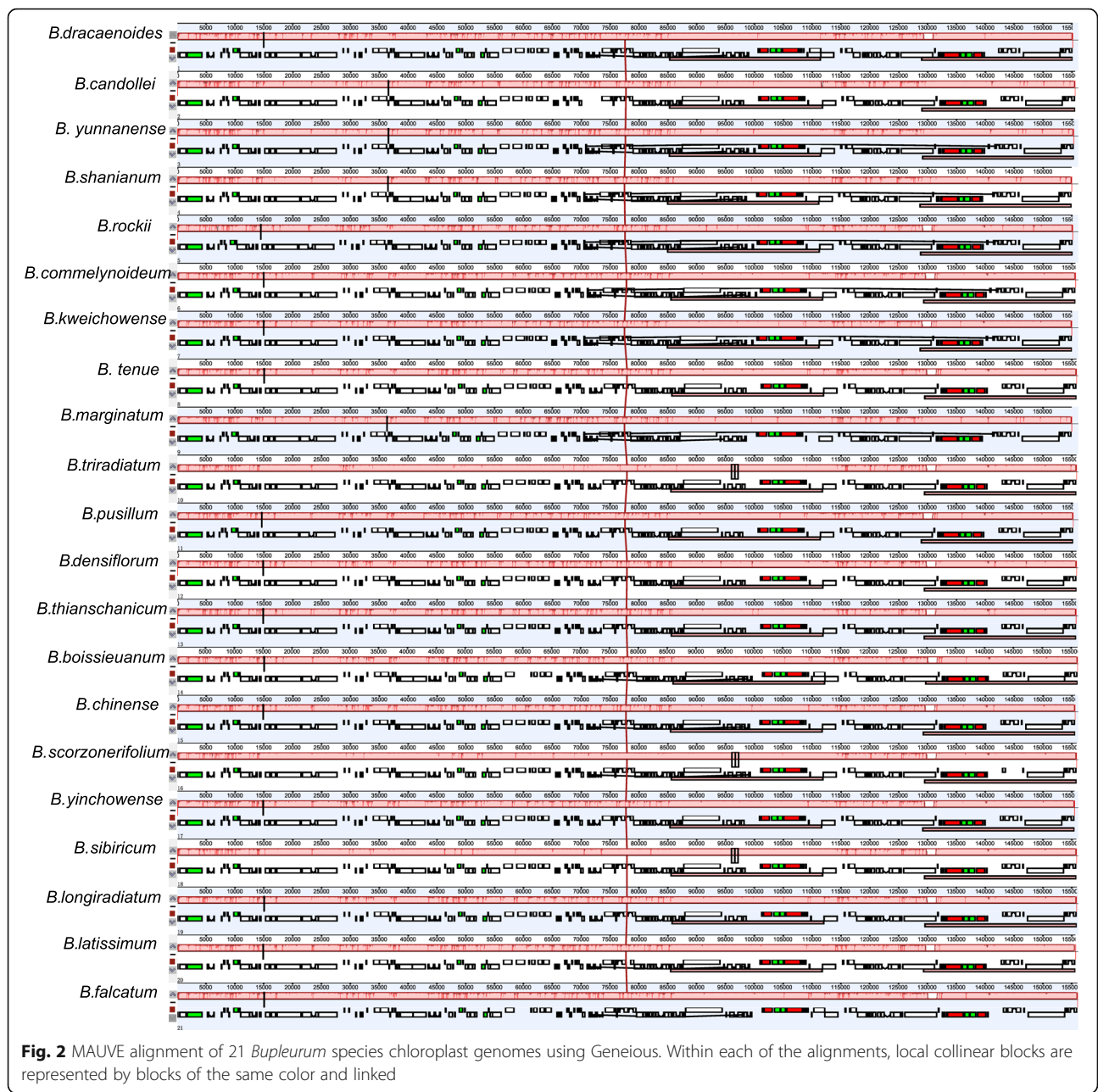
number of SDRs ranged from 34 (*B. candollei* and *B. shanianum*) to 49 (*B. longiradiatum*). The species of group I showed less SDRs than those of group II and III (Additional file 1: Fig. S3). The 21 *Bupleurum* species tended to generate more forward and palindromic repeats rather than reverse repeats, and lacked complement repeats (Additional file 1: Fig. S3).

#### Genome divergent hotspot regions in *Bupleurum* species

Comparative sequence analysis of the 21 *Bupleurum* species using mVISTA revealed a high sequence identity across the 21 species (Additional file 1: Fig. S4), with identity ranged from 96.4 to 99.2%, suggesting that *Bupleurum* chloroplast genomes were fairly conserved. Overall, the coding regions (identity = 99.7 ± 0.8%) were less divergent than non-coding regions (identity: = 96.7 ± 4.1%), and the IR regions (identity = 99.5 ± 1.1%) were more conserved than LSC (identity = 97.1 ± 3.1%) and SSC (identity = 98.2 ± 2.4%) regions. Variations were

observed in some intergenic spacers, including *trnK-rps16*, *rps16-psbK*, *trnG-trnR*, *atpI-rps2*, *trnC-trnT*, *petA-psbJ*, *psaC-ndhG* and *ycf1-trnR*. A few divergent regions were also observed in some coding regions including *psbD*, *atpB*, *ndhD* and *ycf1*.

The nucleotide diversity (Pi) of the chloroplast genomes in the 21 *Bupleurum* species was also calculated to assess the sequence divergence level of these species. In the LSC region, Pi values averaged 0.01087 (range 0.00063–0.03092), and in the SSC region, the average value was 0.01527 (range 0.00394–0.03107); the IR region had the lowest average value of 0.00215 (range 0–0.00963) (Additional file 1: Fig. S5). Most of the sequences with high Pi values were spacer regions between genes. Among these spacer regions, *trnK-rps16*, *rps16-psbK*, *trnG-trnR*, *atpI-rps2*, *trnC-trnT*, *petA-psbJ*, *psaC-ndhG* and *trnR-ycf1* were the 8 regions having Pi values > 0.02000. Only 4 coding regions (*psbD*, *atpD*, *ndhD* and *ycf1*) had high Pi values over 0.02000.

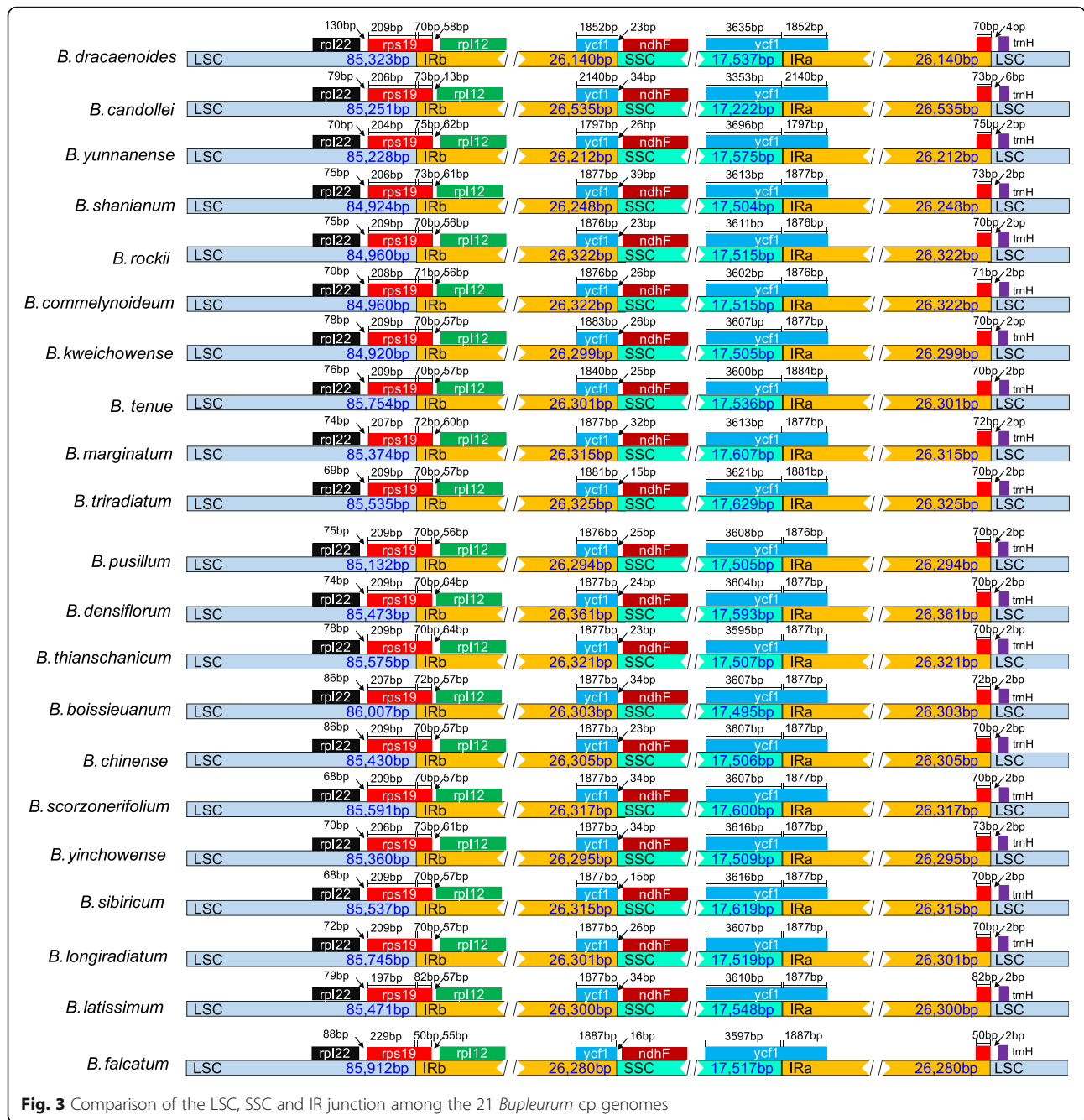


**Selective pressure analysis**

The rate of synonymous substitutions and non-synonymous substitutions (Ka/Ks) of 80 protein-coding genes were calculated to assess the selection pressure between *Bupleurum* species. At the species level, by concatenating all of the 80 genes into a super-matrix, the Ka/Ks ratios ranged from 0.50 (*B. commelynoideum* vs *B. pusillum*) to 5.0 (*B. longiradiatum* vs *B. boissieuana*), with an average ratio of 0.92 (Fig. 4). The Ka/Ks ratios between group II and group III were the highest (averaging 1.10) (Additional file 2: Table S2). The Ka/Ks ratios within group II (averaging 0.98) and group III

(averaging 0.92) were higher than those between group I and group II, (averaging 0.83), between group I and group III (averaging 0.71), and within group I (averaging 0.69) (Additional file 2: Table S2).

The Ka/Ks ratios were also calculated for all the 80 protein-coding genes in the chloroplast genomes of the 21 *Bupleurum* species separately (Additional file 2: Fig. S6 and Fig. S7). Among the genes, *matK* had highest Ka/Ks ratios (around 1.0, especially in group II), following by *ycf2*, *accD*, and *clpP*. The remainder had Ka/Ks ratios ranged from 0 to 0.6. The mean Ka/Ks ratios of protein-coding genes in LSC (0.10) and SSC regions (0.15) were



lower than those in the IR regions (0.19). After sorting the genes into functional categories and groups, Ka/Ks of the photosynthetic genes ( $0.0412 \pm 0.0683$ ) were lower than those of genes related to self-replication ( $0.2507 \pm 0.2197$ ) as well as other genes ( $0.2065 \pm 0.1812$ ).

**Phylogenetic analysis**

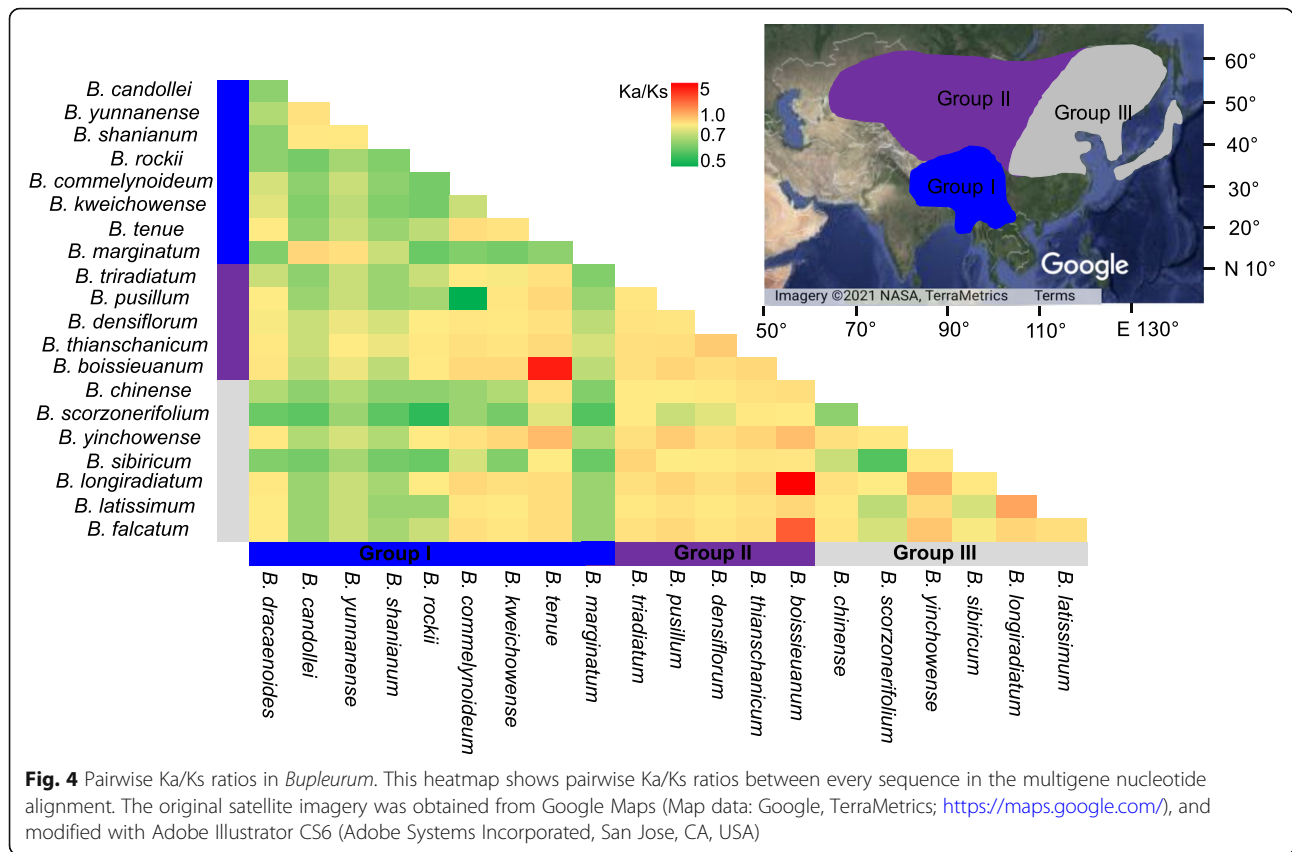
Results of Bayesian and ML analyses of the 21 *Bupleurum* and 2 outgroup chloroplast genomes are presented in Fig. 5. The phylogenetic trees estimated from the Bayesian and ML analyses showed congruence in

their topologies, high bootstrap support values ( $BS > 90\%$ ) and strong posterior probabilities ( $PP = 1$ ) for most of the nodes. The two phylogenetic trees highlighted two clades (clade I and II), containing 4 ( $PP = 100$ ,  $BS = 100\%$ ) and 17 species ( $PP = 100$ ,  $BS = 100\%$ ), respectively.

**Discussion**

**Differences in gene and structure of chloroplast genomes among *Bupleurum* species**

We for the first time described the chloroplast genomes of the 4 *Bupleurum* species endemic to Southwest



**Fig. 4** Pairwise Ka/Ks ratios in *Bupleurum*. This heatmap shows pairwise Ka/Ks ratios between every sequence in the multigene nucleotide alignment. The original satellite imagery was obtained from Google Maps (Map data: Google, TerraMetrics; <https://maps.google.com/>), and modified with Adobe Illustrator CS6 (Adobe Systems Incorporated, San Jose, CA, USA)

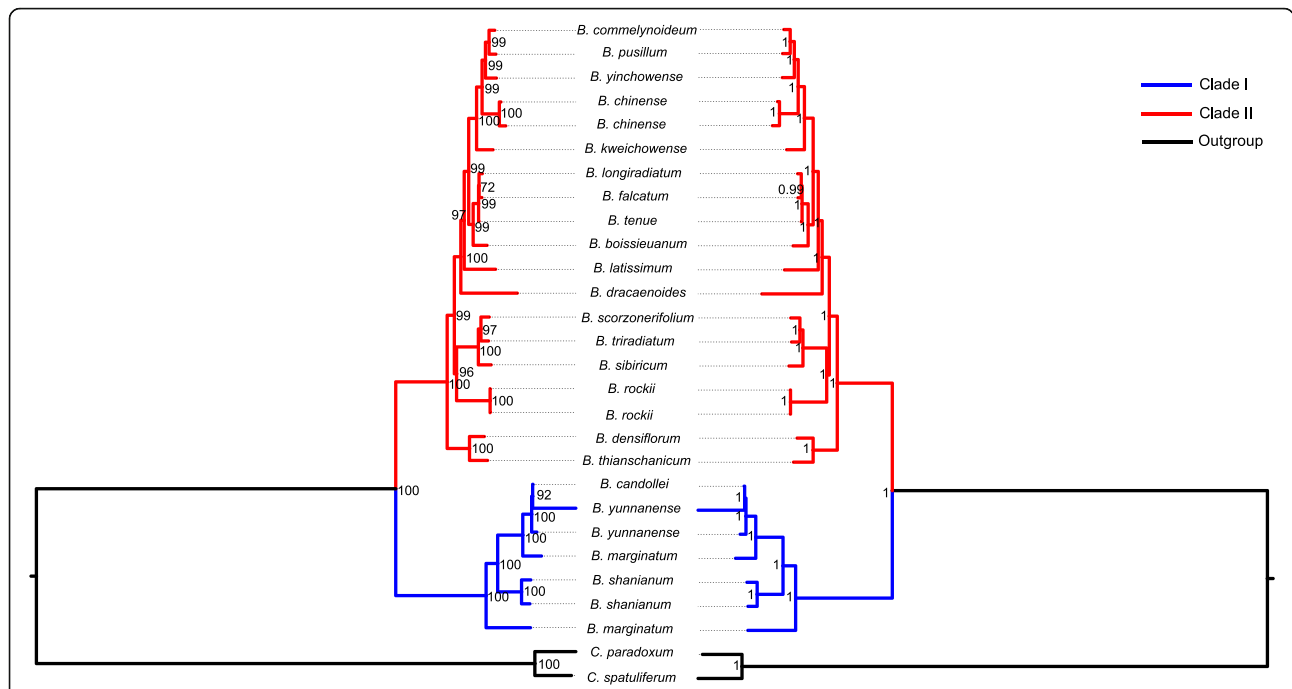
China, which provide important insights into the characteristics of plastid genomes of the members of this genus. Previous studies have suggested that the *Bupleurum* is a monophyletic group based on morphology and molecular (nrDNA and chloroplast fragments) evident. In present study, we provide new insight into the evolution of *Bupleurum* in term of chloroplast genomes [1, 4, 7–14]. Our results showed that the chloroplast genomes of *Bupleurum* species are extremely similar, indicating their high level of conservation especially in terms of chloroplast genome organization, tendency of SDRs and SSRs, gene content. The results support that *Bupleurum* is a monophyletic group in the aspect of chloroplast genomes as they are remarkably conserved.

The chloroplast genomes of the 21 *Bupleurum* species displayed the typical quadripartite structure, comprising a pair of IR regions which were separated by a LSC and a SSC region. The chloroplast genomes size of *Bupleurum*, ranging from 154,925 to 156,108 bp, is larger than those of the other genera in the Apiaceae family (e. g. *Angelica*, *Arracacia*, *Coriandrum*, *Glehnia*, *Heraclium*, *Ligusticum*, *Ostericum*, *Pastinaca*, *Pimpinella*, *Saposhnikovia*, *Semenovia*, *Seseli*, *Tetrataenium* [45–48]). The chloroplast genomes of the *Bupleurum* species

showed only minor differences (~ 1 kb) in sizes. Previous studies suggested that the size variations of the chloroplast genomes resulted from expansion and contraction of the IR regions [20, 24, 48–50]. The IR boundary comparative analysis revealed that gene distributions at SC/IR junctions of *Bupleurum* chloroplast genomes were almost identical, and only minor differences were found in length of these genes (*rps19* and *ycf1*) and SC/IR borders. However, the LSC/IR borders showed differences among *Bupleurum* species and some Apiaceae taxa. For instance, while the LSC/IRb junction was located within the *rps19* gene in *Bupleurum* species as well as some taxa in Apiaceae (*Anthriscus*, *Daucus*, *Tiedemannia*, [48]), it resided within the *rps12* gene in *Anethum* [48], *Foeniculum* [48], *Prangos* [51, 52], *Petroselinum* [48, 52]. Our repetitive sequences analysis in addition to previous studies [32–40] showed that the tendency of SDRs and SSRs were similar among *Bupleurum* species, while the certain diversity in numbers varied.

The gene/intron content and relative gene positions were highly conserved in *Bupleurum* species and almost identical to those in other members of the Apiaceae family [48, 51–56]. Two genes were found to be pseudogenes in *Bupleurum* species. The genes *ycf1* and *rps19*, located in the IRa/SSC and LSC/IRb boundaries,





**Fig. 5** Molecular phylogenetic trees of 21 *Bupleurum* and 2 *Chamaesium* constructed by Maximum likelihood (left) and Bayesian inference (right) analysis based on 80 coding genes. Bootstrap support values (> 50%) based on maximum-likelihood (ML) and posterior probabilities analysis are labelled at each node

respectively, were identified as pseudogenes on account of an incomplete duplication of the normal functional copy.

**High variable regions for potential molecular markers**

Most *Bupleurum* species have important pharmaceutical values and their accurate identification is crucial to their utilization. However, the morphological similarities of the *Bupleurum* species make their authentication extremely difficult. With the development of the molecular technology, many of the chloroplast genome regions, especially the mutational hotspots (e.g. *ndhF*, *matK*, *trnS-trnG*), have been widely implemented [21, 22, 57], while none of the commonly used region in chloroplast genome were effective for identification for different plant taxa.

In a comparative analysis of the complete plastid genome of 5 *Bupleurum* species, only 8 highly variable regions ( $P_i > 0.01$ ) (e. g. *petN-psbM*, *rbcl-accD* and *ccsA-ndhD*) were identified, and all of them were spacer regions between genes [37]. In the present study, we obtained a different result with the supplement of another 17 *Bupleurum* species. The results of mVISTA and slide window analysis showed that the IR regions were more conservative and less variable than the SC regions, possibly as a result of copy number differences in the inverted repeats sequences caused by gene conversion [58]. Different from the findings by Li et al. [37], we identified 4 coding regions

(*psbD*, *atpD*, *ndhD* and *ycf1*) with high  $P_i$  values ( $> 0.02$ ) and the 17 regions reported by Li et al. [37] were not found to be the most variable regions in this study. Instead, another 8 intergenic regions (*trnK-rps16*, *rps16-psbK*, *trnG-trnR*, *atp1-rps2*, *trnC-trnT*, *petA-psbJ*, *psaC-ndhG* and *trnR-ycf1*) were identified, which contained frequent interspecies mutation ( $P_i > 0.02$ ). As more species were included in our analysis, we propose that the 4 genes and 8 intergenic regions we identified can have much potentially to serve as high-resolution DNA barcodes in species authentication of *Bupleurum*.

**Do the genes of *Bupleurum* suffer positive selection under different habitats?**

Plants are subjected to different selection pressures due to different types of stresses in varied habitats, and genes related to a specific environment are usually assumed to be under positive selection [59]. The *Bupleurum* species growing at different latitudes and different altitudes are exposed to different light intensities and temperature, and positive selection is likely to occur among these species in different regions.

Our results indicated that *Bupleurum* species under different habitats suffer positive selection, especially in Northeast China and Northwest China. We found that some genes, including *matK*, *ycf2*, *accD*, and *clpP*, were subjected to positive selection, suggesting that adaptive changes may have occurred more frequently in response

to the highly selective conditions in different habitats. These genes have also previously been found under positive selection [60–62]. The *clpP* gene is essential for plant cells and encodes *clpP* proteases that degrade polypeptides [63]. The *matK* gene is one of the fastest evolutionary genes, which functions in light-regulated activities and plant development [64, 65]. The *accD* gene encodes the  $\beta$ -carboxyl transferase subunit of Acetyl-CoA carboxylase that is essential for the synthesis of products required for the extraplastidic processes needed for leaf development [66]. Gene *ycf2* encodes products that are essential for cell survival [67] and chloroplast protein import [68]. These positively selected genes may contribute to the adaptation of species in *Bupleurum* to various environments. However, our analyses also showed strong purifying selection on most of the genes. Previous studies suggest that purifying selection acting on the genes generally leads to low synonymous and non-synonymous DNA substitution rates in chloroplast genomes of angiosperms, such as Araceae [69, 70] and Liliaceae [71, 72]. Purifying selection, one of the most prevalent mechanisms in natural selection, constantly eliminates deleterious mutations [73]. Most of the genes in these chloroplast genomes were thus subjected to extensive purifying selection to retain conserved functions in *Bupleurum*. The distribution of Ka/Ks indicated that most of the genes in the SSC region have experienced higher selection pressures than those in other chloroplast genome regions, whereas the IR region is more conserved. In addition, the genes involved in photosynthesis tend to have lower rates of evolution than genes related to self-replication and other functions. These differences are likely the results of variations in generation time, gene expression level, gene function, lengths of the encode protein products, and relaxed selection [74–77].

#### Phylogenetic inference from plastid genomes of the genus *Bupleurum* in East Asia

Currently there is no widely acceptable infrageneric classification system of *Bupleurum* [78–80]. East Asian, especially China, is one of the diversity center for the genus *Bupleurum*. *Bupleurum* species in these regions are diverse, exhibiting various life forms (including herbs and shrubs) and pollen types (e. g. subrhomboid, subspheroid and subellipsoid) [40, 81]. However, there are gaps in phylogenetic relationships of East Asian *Bupleurum*, only a few efforts have been made based on morphology [81] and several DNA fragments [4, 13, 14]. Shu et al. proposed to divide Chinese *Bupleurum* into monotypic subgenus *Longifolia* (Wolff) Yuan and subgenus *Eubupleura* (including section *Falcata* and section *Ranunculoidea*) [81], while Wang et al. [4] did not support this treatment and proposed to divide the genus *Bupleurum* into two clades corresponding to the two

subgenera (subg. *Penninervia*, subg. *Bupleurum*) of Neves and Watson [1]. The 21 *Bupleurum* species here have various life forms and widely distribute throughout East Asia (Southwest China; Northwest China; Northeast China, Korea and Japan), providing an opportunity to study the phylogenomic relationship in East Asian *Bupleurum* species. Based on the results of phylogenetic analysis, we propose to divide the 21 *Bupleurum* species, which were all from Asia, into two clades (clade I and clade II). The phylogenetic relationships within *Bupleurum* based on chloroplast genomes we presented herein are largely congruent to that of Wang et al. [4, 13] with only a few differences. *B. rockii* occurred in clade I in our study, but in another clade by Wang et al. [3, 13]. The two clades are likely to correspond to their geographical distribution, not the characteristics of bracteoles. The clade I consists of 4 species from Southwest China, with two endemic species to Southwest China (*B. shanianum* and *B. yunnanense*) and two species (*B. marginatum* and *B. candollei*) with extended distribution in such Himalayan countries as Bhutan, India, Kashmir, Myanmar, Nepal, Pakistan and Sikkim. The species in the clade II, except for *B. dracaenoides*, *B. rockii* and *B. kweichowense*, occur at higher latitudes of China such as Northwest and North China, Japan and Korea. Our results do not fully support Shu et al.'s treatment on *Bupleurum* [81], which divided the subg. *Eubupleura* into sect. *Ranunculoidea* and sect. *Falcata* based on the difference of bracteoles characteristic. However, the subgenus subdivision of the genus *Bupleurum* remains unresolved in this present study because of the lack of chloroplast genome data from subg. *Penninervia*, most species of which are distributed in the Mediterranean region. To access a complete reassessment of the interspecific relationship of *Bupleurum*, more complete plastid genomes information is required, especially the data of subg. *Penninervia*.

#### Conclusion

In this study, we for the first time described 7 full chloroplast genomes of 4 *Bupleurum* species endemic to Southwestern China. Comparative analysis of chloroplast genomes of the 4 species against the published data of 17 *Bupleurum* species revealed that the chloroplast genomes of *Bupleurum* species were extremely conserved with similarities in terms of chloroplast genome organization, tendency of SDRs and SSRs, and gene content. In spite of the highly conservation of the chloroplast genomes of the *Bupleurum* species, some mutational hotspots have been detected, which can be potentially used as high-resolution DNA barcodes in discrimination of *Bupleurum* taxa. *Bupleurum* species under different habitats suffer positive selection, and some genes (*matK*, *ycf2*, *accD*, and *clpP*) are also

subjected to positive selection. Phylogenetic analysis revealed that the 21 *Bupleurum* formed two clades, which are likely to correspond to their geographical distribution. The chloroplast genomes information reported herein and the comparative analysis of *Bupleurum* chloroplast genomes provide important insights into the evolution of the chloroplast genomes and phylogeny of *Bupleurum*.

## Methods

### Sample collection

Leaf from a total of 7 individuals of the 4 *Bupleurum* species (two individuals for each of *B. shaniamum*, *B. yunnanense*, *B. rockii*, while one for *B. kweichowense*) were collected from the wild and were dried with silica gel, and stored at  $-80^{\circ}\text{C}$  until required for DNA extraction. Voucher specimens were collected for each samples and deposited at the herbarium of Southern Medical University. All species were authenticated by prof. Zhi Chao (Southern Medical University). Details information of the samples and voucher numbers of the specimens were shown in Additional file 2: Table S3.

### DNA extraction, library construction and sequencing

Total DNA was extracted from silica-dried leaf material using a modified extraction method described by Yang et al. [82]. The quality and concentration of the extracted DNA were detected by 1.0% agarose gel electrophoresis and by a NanoDrop 2000C spectrophotometry (Thermo Fisher, US). The extraction genomic DNA (approximately 1  $\mu\text{g}$ ) was subjected to random degradation by Covaris (E210), and then fragments with a size of 200–400 bp were selected by using Agencourt AMPure XP-Medium kit. The selected fragments were amplified after suffering from end repair, 3'-adenylation and adaptor ligation. Then, they were heat denatured to single strand after purification. The single strands were circularized, and single strand circle DNA was obtained as the final library. The final library was sequenced by BGISEQ-500 (BGI, Shenzhen, China) to generate raw reads. The details of the quantity and quality of raw reads, and coverage depth of the assembled genomes were provided in Additional file 2: Table S4.

### Genome assembly and annotation

The generated raw sequencing data was filtered using program SOAPnuke [83] with default parameters to remove adapters, low-quality reads with quality value  $\leq 10$ , to final obtain high-quality reads. Then, the high-quality reads were aligned to the published *B. commelynoideum* chloroplast genome (NCBI accession MT162552) using Geneious v 10.2.2 [84] with default settings. Subsequently, the matched reads were selected for de novo assembled with SPAdes v3.11.1 [85]. The accuracy of

assembly was evaluated by detecting the sequence coverage and the reading segment coverage at the contig connection.

The assembled chloroplast genome annotations were annotated using GeSeq [86] with the reference chloroplast genome of *B. commelynoideum* (NCBI accession MT162552). All the tRNAs were scanned with tRNAscan-SE [87] and ARAGORN [88]. The visual presentations of the physical circular maps of the genomes were generated using OGDRAW [89]. Finally, the annotated chloroplast genomes of the 4 *Bupleurum* species were submitted to the National Center for Biotechnology Information database (NCBI) under the accession numbers MW135450–MW135456, which were listed in Additional file 2: Table S3.

### Genome structure and comparative analysis

First, the chloroplast genome characteristics of the 4 *Bupleurum* species were described. Therein, Geneious R8.1 [90] was employed to conduct the GC content. MISA-web (<https://webblast.ipk-gatersleben.de/misa/>) was implemented to search simple sequence repeats (SSRs) with minimum numbers of 10, 5, 4, 3, 3, 3 repeat units for mono-, di-, tri- tetra-, penta-, and hexanucleotide SSRs, respectively. Short dispersed repeats (SDRs) analysis was implemented in REPuter [91] with the following parameters: a minimal repeat size of 30 bp, and sequence identity  $\geq 90\%$  (hamming distance of 3 kb). To determine the IR expansion/contraction, genes distributed in and beside the borders of LSC, SSC and IR regions were compared.

In order to examine the divergence hotspots among the *Bupleurum* species, the whole chloroplast sequences of the 4 *Bupleurum* species, together with 17 published chloroplast genomes of *Bupleurum* species downloaded from NCBI database (Additional file 2: Table S5), were aligned using Geneious software. Subsequently, they were compared and visualized using mVISTA (<http://genome.lbl.gov/vista/mvista/submit.shtml>) with the reference chloroplast genome sequence of *B. commelynoideum* (NCBI accession MT162552). DnaSP v. 6.0 was used for sliding window analysis for computing nucleotide diversity ( $\pi$ ) among the chloroplast genome sequences [92], with 600 bp windows size and 200 bp step size.

Selective pressure estimation for the 21 *Bupleurum* species was carried out by calculating the ratio of the non-synonymous substitution ( $K_a$ ) to the synonymous substitution rate ( $K_s$ ) for all protein-coding genes sequences in DnaSP v6.

### Phylogenetic analysis

In order to gain insight into the phylogeny of East Asian *Bupleurum*, a total of 19 available chloroplast complete

genomes of 17 *Bupleurum* species were downloaded from the NCBI database (Additional file 2: Table S4). The chloroplast genomes data of two *Chamaesium* species (*C. paradoxum*, NCBI accession MK780227; *C. spatuliferum*, NCBI accession MN119371), belonging to subfamily Apioideae, were also downloaded and set as outgroups (Additional file 2: Table S4). Only a dataset of 80 protein-coding genes (PCGs) was used for the phylogenetic analyses. A 23-taxon sequence matrix including two outgroups were aligned using the Geneious software.

Phylogeny was conducted through two approaches, namely the Maximum likelihood (ML) analyses and a Bayesian inference (BI) analyses. ML phylogenetic analysis was performed in RAxML v8.2.4 [93]. First, the best likelihood tree was obtained from 100 starting trees using rapid bootstrap analyses with 1000 replicates. Multiparametric bootstrapping analyses with 1000 replicates was conducted to obtain the bootstrap for each node. Substitution model for the two analyses were GTRGAMMA model. Bayesian inference (BI) was conducted in MrBayes v3.2.6 [94, 95]. The best-fit nucleotide substitution model (GTR + I + GAMMA) for Bayesian analysis was inferred from jModelTest v. 2.1.10 under the Akaike information criteria (AIC) [96]. Markov chain Monte Carlo (MCMC) analysis was performed with 50 million generations and sampling trees every 5000 generations. The first 10% of trees were discarded as burn fraction, and the remaining trees were combined to synthesize the consistent tree and estimate posterior probabilities. The resulting trees were rooted with *C. paradoxum* and *C. spatuliferum* and visualized with FigTree v 1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

#### Abbreviations

ITS: Internal transcribed spacer region; nrDNA: Nuclear ribosome DNA; NGS: Next-generation sequencing; Cp: Chloroplast; LSC: Large single copy; SSC: Small single copy; IR: Inverted repeat; tRNA: Transfer RNA; rRNA: Ribosomal RNA; SSR: Simple sequence repeat; SDR: Short dispersed repeat; Met: Methionine; Trp: Tryptophan; Arg: Arginine; Leu: Leucine; Ser: Serine; RSCU: Synonymous codon usage; GC3: GC content on the third synonymous codon position; Ks: Rate of synonymous substitutions; Ka: Rate of non-synonymous substitutions; Pi: Polymorphic information; ML: Maximum likelihood; BS: Bootstrap support value; PP: Posterior probability; BI: Bayesian inference; AIC: Akaike information criteria; MCMC: Markov chain Monte Carlo.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-08008-z>.

**Additional file 1.**

**Additional file 2.**

#### Acknowledgements

Not applicable.

#### Authors' contributions

ZC conceived the study and provided the funding, reviewed and revised the drafts of the paper. RH analyzed data, and wrote and revised the manuscript. XNX performed the genome assembly and annotation, and assisted with the molecular experiments and data analysis. AMC conducted the molecular experiments, performed the genome assembly and annotation. FL assisted with the molecular experiments and the genome assembly. EWT collected field samples and assisted with the molecular experiments. All authors read and approved the final manuscript.

#### Funding

The work was financially supported by National Natural Science Foundation of China (No. 30650652, 81373905), Guangdong Provincial Natural Science Foundation (No. 2014A030313321) and China Postdoctoral Science Foundation (No. 2021 M691482). The funders had no role in study design, data collection, analysis and interpretation, or preparation of the manuscript.

#### Availability of data and materials

Sequence information of the 7 chloroplast genomes of the four *Bupleurum* species (*B. shanianum*, *B. yunnanense*, *B. kweichowense* and *B. rockii*) is available in the National Center for Biotechnology Information database (NCBI) under the accession number MW135450-MW135456 (Additional file 2: Table S3). The accession numbers corresponding to the additional datasets used and analysed in this study can be found in Additional file 2: Table S5. These additional data were retrieved from the National Center for Biotechnology Information database (<https://www.ncbi.nlm.nih.gov/>).

#### Declarations

##### Ethics approval and consent to participate

No specific permissions were required for the collection of plant material that was conducted in this study. Our field works and molecular experiments were carried out in compliance with the relevant laws of China.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Department of Pharmacy, Zhujiang Hospital, Southern Medical University, Guangzhou 510282, China. <sup>2</sup>Faculty of Medicinal Plants and Pharmacognosy, School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, China. <sup>3</sup>Guangdong Provincial Key Laboratory of Chinese Medicine Pharmaceuticals, Guangzhou 510515, China.

Received: 25 October 2020 Accepted: 13 September 2021

Published online: 02 October 2021

#### References

1. Neves SS, Watson MF. Phylogenetic relationships in *Bupleurum* (Apiaceae) based on nuclear ribosomal DNA ITS sequence data. *Ann Bot*. 2004;93(4): 379–98. <https://doi.org/10.1093/aob/mch052>.
2. Sheh ML, Watson MF. *Bupleurum* L. In: Flora of China Editorial Committee. Beijing: Flora of China, vol. 14. Science Press and Saint Louis: Missouri Botanical Garden Press; 2005. p. 60–74.
3. Stepanova AV, Oskolski AA. Wood anatomy of *Bupleurum* L. (Apiaceae, Apiaceae) in relation to habit, phylogenetic relationships, and infrageneric taxonomy. *Plant Div Evol*. 2010;128(3-4):501–16. <https://doi.org/10.1127/1869-6155/2010/0128-0024>.
4. Wang BC, Ma XG, He XJ. A taxonomic re-assessment in the Chinese *Bupleurum* (Apiaceae): insights from morphology, nuclear ribosomal internal transcribed spacer, and chloroplast (*trnH-psbA*, *matK*) sequences. *J Syst Evol*. 2011;49(6):558–89. <https://doi.org/10.1111/j.1759-6831.2011.00157.x>.
5. Ma XG, Zhao C, Liang QL, He XJ. *Bupleurum baimaense* (Apiaceae), a new species from Hengduan Mountains, China. *Ann Bot Fennici*. 2013;50(6):379–85. <https://doi.org/10.5735/085.050.0601>.
6. Ma XG, He XJ. *Bupleurum shanianum* sp. nov. (Apiaceae) from China. *Nord J Bot*. 2015;33:71–3.

7. Downie SR, Katz-Downie DS, Watson MF. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rp16* and *rpoC1* intron sequences: towards a suprageneric classification of subfamily Apioideae. *Am J Bot.* 2000;87(2):273–92. <https://doi.org/10.2307/2656915>.
8. Downie SR, Watson MF, Spalik K, Katz-Downie DS. Molecular systematics of Old World Apioideae (Apiaceae): relationships among some members of tribe Peucedaneae sensu lato, the placement of several island-endemic species, and resolution within the apioid superclade. *Can J Bot.* 2000;78:506–28.
9. Calviño C, Tilney P, Van Wyk BE, Downie SR. A molecular phylogenetic study of southern African Apiaceae. *Am J Bot.* 2006;93(12):1828–47. <https://doi.org/10.3732/ajb.93.12.1828>.
10. Zhou J, Gong X, Downie SR, Peng H. Towards a more robust molecular phylogeny of Chinese Apiaceae subfamily Apioideae: additional evidence from nrDNA ITS and cpDNA intron (*rp16* and *rps16*) sequences. *Mol Phylogenet Evol.* 2009;53(1):56–68. <https://doi.org/10.1016/j.ympev.2009.05.029>.
11. Downie SR, Spalik K, Katz-Downie DS, Reduron JP. Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. *Plant Divers Evol.* 2010;128(1-2):111–36. <https://doi.org/10.1127/1869-6155/2010/0128-0005>.
12. Wen J, Yu Y, Xie DF, Peng C, Liu Q, Zhou SD, et al. A transcriptome-based study on the phylogeny and evolution of the taxonomically controversial subfamily Apioideae (Apiaceae). *Ann Bot.* 2020;125(6):937–53. <https://doi.org/10.1093/aob/mcaa011>.
13. Wang QZ, He XJ, Zhou SD, Wu YK, Yu Y, Pang YL. Phylogenetic inference of the genus *Bupleurum* (Apiaceae) in Hengduan Mountains based on chromosome counts and nuclear ribosomal DNA ITS sequences. *J Syst Evol.* 2008;46:142–54.
14. Wang QZ, Zhou SD, Liu TY, Pang YL, Wu YK, He XJ. Phylogeny and classification of the Chinese *Bupleurum* (Apiaceae) based on nuclear ribosomal DNA internal transcribed spacer and *rps16*. *Acta Biol Cracov Ser Bot.* 2008;50:105–16.
15. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA.* 1987;84(24):9054–8. <https://doi.org/10.1073/pnas.84.24.9054>.
16. Ravi V, Khurana J, Tyagi A, Khurana P. An update on chloroplast genomes. *Plant Syst Evol.* 2008;271(1-2):101–22. <https://doi.org/10.1007/s00606-007-0608-0>.
17. Dempewolf H, Kane NC, Ostevik KL, Geleta M, Barker MS, Lai Z, et al. Establishing genomic tools and resources for *Guizotia abyssinica* (f) Cass-the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Mol Ecol Res.* 2010;10:1048–58.
18. Huang YL, Li XJ, Yang ZY, Yang CJ, Yang JB, Ji YH. Analysis of complete chloroplast genome sequences improves phylogenetic resolution in *Paris* (Melanthiaceae). *Front Plant Sci.* 2016;7:1797. <https://doi.org/10.3389/fpls.2016.01797>.
19. Xue S, Shi T, Luo W, Ni X, Iqbal S, Ni Z, et al. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hortic Res.* 2019;6:89.
20. Duan H, Guo JB, Xuan L, Wang ZY, Li MZ, Yin YL, et al. Comparative chloroplast genomics of the genus *Taxodium*. *BMC Genomics.* 2020;21(1):114. <https://doi.org/10.1186/s12864-020-6532-1>.
21. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci USA.* 2010;107(10):4623–8. <https://doi.org/10.1073/pnas.0907801107>.
22. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot.* 2014; 101(11):1987–2004. <https://doi.org/10.3732/ajb.1400398>.
23. Sun Y, Moore MJ, Zhang S, Soltis PS, Soltis DE, Zhao T, et al. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. *Mol Phylogenet Evol.* 2016;96:93–101. <https://doi.org/10.1016/j.ympev.2015.12.006>.
24. Wicke S, Naumann J. Molecular evolution of plastid genomes in parasitic flowering plants. *Adv Bot Res.* 2017;85:312–47.
25. Yan M, Fritsch PW, Moore MJ, Feng T, Meng A, Yang J, et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. *Mol Phylogenet Evol.* 2018;121:198–211. <https://doi.org/10.1016/j.ympev.2018.01.004>.
26. Saarela JM, Wysocki WP, Barrett CF, Soreng RJ, Davis JI, Clark LG, et al. Plastid phylogenomics of the cool-season grass subfamily: clarification of relationships among early-diverging tribes. *AoB Plants.* 2015;7:1–27. <https://doi.org/10.1093/aobpla/plv046>.
27. Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. Plastome phylogenetics: 30 years of inferences into plant evolution. In: Chaw SM, Jansen RK, editors. *Advances in botanical research*. London: Academic; 2018. p. 293–313. <https://doi.org/10.1016/bs.abr.2017.11.016>.
28. Heckenbauer J, Paun O, Chase MW, Ashton PS, Kamariah AS, Samuel R. Molecular phylogenomics of the tribe Shoreaeae (Dipterocarpaceae) using whole plastid genomes. *Ann Bot.* 2019;123(5):857–65. <https://doi.org/10.1093/aob/mcy220>.
29. Thode VA, Lohmann LG. Comparative chloroplast genomics at low taxonomic levels: a case study using *Amphilophium* (Bignoniaceae, Bignoniaceae). *Front Plant Sci.* 2019;10:796. <https://doi.org/10.3389/fpls.2019.00796>.
30. Liu E, Yang C, Liu J, Jin S, Harijati N, Hu Z, et al. Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species. *Sci Rep.* 2019;9(1):809. <https://doi.org/10.1038/s41598-018-37456-z>.
31. Ren T, Li ZX, Xie DF, Gui LJ, Peng C, Wen J, et al. Plastomes of eight *Ligusticum* species: characterization, genome evolution, and phylogenetic relationships. *BMC Plant Biol.* 2020;20(1):519. <https://doi.org/10.1186/s12870-020-02696-7>.
32. Shin DH, Lee JH, Kang SH, Ahn BO, Kim CK. The complete chloroplast genome of the hare's ear root, *Bupleurum falcatum*: Its molecular features. *Genes.* 2016;7(5):20. <https://doi.org/10.3390/genes7050020>.
33. Wu Y, Zhang TZ, Qiu DY, Chai Q, Fan WB, Li ZH, et al. Complete plastid genome of *Bupleurum boissieuianum*, an endemic herb plant in western China. *Conserv Genet Resour.* 2018;10(4):635–7. <https://doi.org/10.1007/s12686-017-0890-2>.
34. Zhang F, Zhao ZY, Yuan QJ, Chen SQ, Huang LQ. The complete chloroplast genome sequence of *Bupleurum chinense* DC. (Apiaceae). *Mitochondrial DNA B.* 2019;4(2):3665–6. <https://doi.org/10.1080/23802359.2019.1678427>.
35. Deng XD, Liu HZ, Yang YH, Liang YH, Luo KF, Yang H. The first complete chloroplast genome sequence of the medicinal plant *Bupleurum marginatum* (Apiaceae). *Mitochondrial DNA B.* 2020;5(2):1836–8. <https://doi.org/10.1080/23802359.2020.1752122>.
36. Kim CK, Jin MW, Kim YK. The complete mitochondrial genome sequences of *Bupleurum falcatum* (Apiales: Apiaceae). *Mitochondrial DNA B.* 2020;5(3):2576–7. <https://doi.org/10.1080/23802359.2020.1781566>.
37. Li J, Xie DF, Guo XL, Zheng ZY, He XJ, Zhou SD. Comparative analysis of the complete plastid genome of five *Bupleurum* species and new insights into DNA barcoding and phylogenetic relationship. *Plants.* 2020;9(4):543. <https://doi.org/10.3390/plants9040543>.
38. Yang L, Xiong F, Xiao Y, Li J, Chen C, Zhou G. The complete chloroplast genome of *Bupleurum longicaule* var. *strictum*, an annual herb endemic to China. *Mitochondrial DNA B.* 2020;5(1):899–901. <https://doi.org/10.1080/23802359.2020.1718024>.
39. Zhang F, Yang Z, Wang Z, Zhao Z, Xie B, Chen S. The complete chloroplast genome sequence of *Bupleurum scorzoniferifolium* Willd. (Apiaceae). *Mitochondrial DNA B.* 2020;5(2):1998–9. <https://doi.org/10.1080/23802359.2020.1756489>.
40. Zhao Z, Liu J, Zhou M, Pan Y. Chloroplast genome characterization of *Bupleurum dracaenoides*, a critically endangered woody species endemic to China, with insights of Apioideae phylogeny. *Gene Rep.* 2020;20:100784. <https://doi.org/10.1016/j.genrep.2020.100784>.
41. He K, Hu NQ, Chen X, Li JT, Jiang XL. Interglacial refugia preserved high genetic diversity of the Chinese mole shrew in the mountains of Southwest China. *Heredity.* 2016;116(1):23–32. <https://doi.org/10.1038/hdy.2015.62>.
42. Lu LM, Mao LF, Yang T, Ye JF, Liu B, Liu B, et al. Evolutionary history of the angiosperm flora of China. *Nature.* 2018;554(7691):234–8. <https://doi.org/10.1038/nature25485>.
43. Lamoreaux JF, Morrison JC, Ricketts TH, Olson DM, Dinerstein E, McKnight MW, et al. Global tests of biodiversity concordance and the importance of endemism. *Nature.* 2006;440(7081):212–4. <https://doi.org/10.1038/nature04291>.
44. Castellanos-Morales G, Gámez N, Castillo-Gámez RA, Eguarte LE. Peripatric speciation of an endemic species driven by Pleistocene climate change: the case of the Mexican prairie dog (*Cynomys mexicanus*). *Mol Phylogenet Evol.* 2016;94(Pt A):171–81. <https://doi.org/10.1016/j.ympev.2015.08.027>.

45. Leitão RP, Zuanon J, Villéger S, Williams SE, Baraloto C, Fortunel C, et al. Rare species contribute disproportionately to the functional structure of species assemblages. *Proc Biol Sci.* 2016;283(1828):1828. <https://doi.org/10.1098/rspb.2016.0084>.
46. Naumenko SA, Logacheva MD, Popova NV, Klepikova AV, Penin AA, Bazykin GA, et al. Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection. *Mol Ecol.* 2017;26(2):536–53. <https://doi.org/10.1111/mec.13927>.
47. Miller CN, Kwit C. Overall seed dispersal effectiveness is lower in endemic Trillium species than in their widespread congeners. *Am J Bot.* 2018;105(11):1847–57. <https://doi.org/10.1002/ajb2.1188>.
48. Samigullin TH, Logacheva MD, Terenteva EI, Degtjareva GV, Vallejo-Roman CM. Plastid genome of *Seseli montanum*: complete sequence and comparison with Plastomes of other members of the Apiaceae Family. *Biochemistry.* 2016;81(9):981–5. <https://doi.org/10.1134/S0006297916090078>.
49. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 2004;11(4):247–61. <https://doi.org/10.1093/dnares/11.4.247>.
50. Fu CN, Li HT, Milne R, Zhang T, Ma PF, Yang J, et al. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. *BMC Genomics.* 2017;18(1):956. <https://doi.org/10.1186/s12864-017-4319-9>.
51. Samigullin TH, Logacheva MD, Degtjareva GV, Terenteva EI, Vallejo-Roman CM. Complete plastid genome of critically endangered plant *Prangos trifida* (Apiaceae: Apioideae). *Conserv Genet Resour.* 2018;10(4):847–9. <https://doi.org/10.1007/s12686-017-0945-4>.
52. Mustafina FU, Yi DK, Choi K, Shin CH, Tojibaev KS, Downie SR. A comparative analysis of complete plastid genomes from *Prangos fedtschenkoi* and *Prangos lipskyi* (Apiaceae). *Ecol Evol.* 2019;9(1):364–77. <https://doi.org/10.1002/ece3.4753>.
53. Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, et al. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics.* 2006;7(1):222. <https://doi.org/10.1186/1471-2164-7-222>.
54. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst Bot.* 2015;40(1):336–51. <https://doi.org/10.1600/036364415X686620>.
55. Tian EW, Liu QQ, Chen WN, Li F, Chen AM, Li C, et al. Characterization of complete chloroplast genome of *Angelica sinensis* (Apiaceae), an endemic medical plant to China. *Mitochondrial DNA Res.* 2019;4(1):158–9. <https://doi.org/10.1080/23802359.2018.1544862>.
56. Zheng HY, Guo XL, He XJ, Yu Y, Zhou SD. The complete chloroplast genome of *Chamaesium paradoxum*. *Mitochondrial DNA Res.* 2019;4(1):2069–70. <https://doi.org/10.1080/23802359.2019.1617064>.
57. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One.* 2012;7(4):1–9. <https://doi.org/10.1371/journal.pone.0035071>.
58. Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 2006;46(1):85–94. <https://doi.org/10.1111/j.1365-313X.2006.02673.x>.
59. Yang Z, Wong WSW, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005;22(4):1107–18. <https://doi.org/10.1093/molbev/msi097>.
60. Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front Plant Sci.* 2016;7:972. <https://doi.org/10.3389/fpls.2016.00959>.
61. Choi KS, Kwak M, Lee B, Park SJ. Complete chloroplast genome of *Tetragonia tetragonioides*: molecular phylogenetic relationships and evolution in Caryophyllales. *PLoS One.* 2018;13(6):1–11. <https://doi.org/10.1371/journal.pone.0199626>.
62. Yu X, Zuo L, Lu D, Lu B, Yang M, Wang J. Comparative analysis of chloroplast genomes of five *Robinia* species: genome comparative and evolution analysis. *Gene.* 2019;689:141–51. <https://doi.org/10.1016/j.gene.2018.12.023>.
63. Kuroda H, Maliga P. The plastid clpP1 protease gene is essential for plant development. *Nature.* 2003;425(6953):86–9. <https://doi.org/10.1038/nature01909>.
64. Barthelet MM, Hilu KW. Expression of *matK*: functional and evolutionary implications. *Am J Bot.* 2007;94(8):1402–12. <https://doi.org/10.3732/ajb.94.8.1402>.
65. Barthelet MM, Hilu KW. Evaluating evolutionary constraint on the rapidly evolving gene *matK* using protein composition. *J Mol Evol.* 2008;66(2):85–97. <https://doi.org/10.1007/s00239-007-9060-6>.
66. Kode V, Mudd EA, Lamtham S, Day A. The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J.* 2005;44(2):237–44. <https://doi.org/10.1111/j.1365-313X.2005.02533.x>.
67. Drescher A, Ruf S, Calsa T, Carrer H, Bock R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 2000;22(2):97104. <https://doi.org/10.1046/j.1365-313x.2000.00722.x>.
68. Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, et al. A Ycf2-FtsHi Heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell.* 2018;30(11):2677–703. <https://doi.org/10.1105/tpc.18.00357>.
69. Henriquez CL, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, et al. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). *Planta.* 2020;251:72.
70. Kim SH, Yang JY, Park J, Yamada T, Maki M, Kim SC. Comparison of whole plastome sequences between thermogenic skunk cabbage *Symplocarpus renifolius* and nonthermogenic *S. nipponicus* (Oronitoidae; Araceae) in East Asia. *Int J Mol Sci.* 2019;20(19):4678. <https://doi.org/10.3390/ijms20194678>.
71. Gao XY, Zhang X, Meng HH, Li J, Zhang D, Liu CN, et al. Comparative chloroplast genomes of Paris sect. *Marmorata*: insights into repeat regions and evolutionary implications. *BMC Genomics.* 2018;19:878.
72. She RX, Zhao P, Zhou HJ, Yue M, Yan F, Hu GJ, et al. Complete chloroplast genomes of Liliaceae (s.l.) species: comparative genomic and phylogenetic analyses. *Nord J Bot.* 2020. <https://doi.org/10.1111/njb.02477>.
73. Wu Z, Liao R, Yang T, Dong X, Lan D, Qin R, et al. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics.* 2020;21(1):621. <https://doi.org/10.1186/s12864-020-07045-4>.
74. Muse SV, Gaut BS. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. *Genetics.* 1997;146(1):393–9. <https://doi.org/10.1093/genetics/146.1.393>.
75. Mcinerney JO. The causes of protein evolutionary rate variation. *Trends Ecol Evol.* 2006;21(5):230–2. <https://doi.org/10.1016/j.tree.2006.03.008>.
76. Wang Z, Zhang J. Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet.* 2009;5(1):e1000329. <https://doi.org/10.1371/journal.pgen.1000329>.
77. Alvarez-Ponce D, Fares MA. Evolutionary rate and duplicability in the Arabidopsis thaliana protein–protein interaction network. *Genome Biol Evol.* 2012;4(12):1263–74. <https://doi.org/10.1093/gbe/evs101>.
78. Tutin TG. *Bupleurum*. In: Tutin TG, editor. *Flora Europaea*, vol. 2. Cambridge: Cambridge University Press; 1968. p. 345–50.
79. Wolff H. *Bupleurum*. In: Engler A, editor. *Das pflanzenreich regni vegetabilis conspectus*. Leipzig: Wilhelm Engelmann; 1980. p. 36–173.
80. Snogerup S, Snogerup B. *Bupleurum* L. (Umbelliferae) in Europe I. The annuals, B. sect. *Bupleurum* and set *Aristata*. *Willdenowia.* 2001;31(2):205–308. <https://doi.org/10.3372/wi.31.31201>.
81. Shu P, Yuan CQ, She ML, Liu YZ, Xiang BR, An DK. Numerical taxonomy of medical *Bupleurum* species in China. *Acta Botan Boreali-Occiden Sin.* 1998;18:277–83.
82. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Res.* 2014;14:1024–31. <https://doi.org/10.1111/1755-0998.12251>.
83. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience.* 2018;7(1):1–6. <https://doi.org/10.1093/gigascience/gix120>.
84. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
85. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
86. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq: versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–11. <https://doi.org/10.1093/nar/gkx391>.

87. Lowe TM, Chan PP. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016;44(W1):W54–7. <https://doi.org/10.1093/nar/gkw413>.
88. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 2004;32(1):11–6. <https://doi.org/10.1093/nar/gkh152>.
89. 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. *Curr Genet.* 2007;52(5–6):26–74. <https://doi.org/10.1007/s00294-007-0161-y>.
90. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199>.
91. 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(22):4633–42. <https://doi.org/10.1093/nar/29.22.4633>.
92. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25(11):1451–2. <https://doi.org/10.1093/bioinformatics/btp187>.
93. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
94. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 2001;17(8):754–5. <https://doi.org/10.1093/bioinformatics/17.8.754>.
95. Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19(12):1572–4. <https://doi.org/10.1093/bioinformatics/btg180>.
96. Darriba D, Taboada GL, Doallo R, Posada D, Europe PMC Funders Group. jModelTest 2: more models, new heuristics and high-performance computing. *Nat Methods.* 2015;9:6–9.

## Publisher's Note

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

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

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

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

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

