BMC Genomics

Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae

Dhafer A. Alzahrani[1], Samaila S. Yaradua[1,2*], Enas J. Albokhari[1,3] and Abidina Abba[1]

Abstract

Background: The plastome of medicinal and endangered species in Kingdom of Saudi Arabia, *Barleria prionitis* was sequenced. The plastome was compared with that of seven Acanthoideae species in order to describe the plastome, spot the microsatellite, assess the dissimilarities within the sampled plastomes and to infer their phylogenetic relationships.

Results: The plastome of *B. prionitis* was 152,217 bp in length with Guanine-Cytosine and Adenine-Thymine content of 38.3 and 61.7% respectively. It is circular and quadripartite in structure and constitute of a large single copy (LSC, 83, 772 bp), small single copy (SSC, 17, 803 bp) and a pair of inverted repeat (IRa and IRb 25, 321 bp each). 131 genes were identified in the plastome out of which 113 are unique and 18 were repeated in IR region. The genome consists of 4 rRNA, 30 tRNA and 80 protein-coding genes. The analysis of long repeat showed all types of repeats were present in the plastome and palindromic has the highest frequency. A total number of 98 SSR were also identified of which mostly were mononucleotide Adenine-Thymine and are located at the non coding regions. Comparative genomic analysis among the plastomes revealed that the pair of the inverted repeat is more conserved than the single copy region. In addition high variation is observed in the intergenic spacer region than the coding region. The genes, *ycf1* and *ndhF* and are located at the border junction of the small single copy region and IRb region of all the plastome. The analysis of sequence divergence in the protein coding genes indicates that the following genes undergo positive selection (*atpF, petD, psbZ, rpl20, petB, rpl16, rps16, rpoC, rps7, rpl32* and *ycf3*). Phylogenetic analysis indicated sister relationship between Ruellieae and Justcieae. In addition, *Barleria, Justicia* and *Ruellia* are paraphyletic, suggesting that Justiceae, Ruellieae, Andrographideae and Barlerieae should be treated as tribes.

Conclusions: This study sequenced and assembled the first plastome of the taxon *Barleria* and reported the basics resources for evolutionary studies of *B. prionitis* and tools for phylogenetic relationship studies within the core Acanthaceae.

Keywords: Acanthoideae, Chloroplast genome, *Barleria prionitis*, Phylogenomics

* Correspondence: dryaradua@gmail.com
[1]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia
[2]Department of Biology, Umaru Musa Yaradua University, Centre for Biodiversity and Conservation, Katsina, Nigeria
Full list of author information is available at the end of the article

Background

The Acanthaceae Juss. Ex Bercht.& J. Presl is among the largest family in the order Lamiales with ca. 3800 recognized species accommodated in ca. 200 genera [1], the members of the family are mainly diversified in the sub tropics and tropics, with few species in the temperate zones [2]. The family is close to Bignoniaceae family in the Lamiales order [3]. The main centres of distribution of the species in the family are Africa, Central America and Asian continent particularly Malaysia, Indonesia and Brazil [4]. They are characterized by having decussate phyllotaxis, while some species have congest whorled phyllotaxis, the leaves are usually simple with toothed margin, opposite, existipulate and contained calcium oxalate crystals or hypodermal calcium carbonate cystolith [5, 6].

In an effort to resolve taxonomic issues of the family and its species, researchers for the past decades works extensively in delimiting the family [7–10], identifying major clades in the family [11–14]. Scotland and his colleagues carried out infrafamilial studies using floral parts [15–17], their findings gives more insight on the infra familial classification of the family and gives morphological synapomorphies of the major lineages. Recently, phylogenetic approach was used to reveal the relationships between the lineages [18–20]. Despite these researches, the classifications of the species within the Acanthoideae are still not clear.

The chloroplast organelle is one the most distinguishing featured that differentiates plant cell and other type of cells; therefore it is the most noticeable feature in plants. The organelle which is semi-autonomous is believed to have evolved decade of millions years ago from cynobacterium [21, 22]. The plastome of flowering plant is conserved than the other genomes (i.e mitochondrial and nuclear genomes), in addition the genome is small compared with the others and it is used frequently in phylogeny studies due to its low rate of nucleotide substitution [23]. The chloroplast genome is typically quadripartite in structure, containing large single copy (LSC) and small single copy (SSC) separated by pair of inverted repeat (IR) [24]. The genome organization, its content and gene structure are highly conserved [25]. Due to its conserved nature, the cp genome contents are widely used by researchers as a tool to investigate phylogenetic relationship and in genomic studies [26]. Single nucleotide polymorphisms as well as insertion/deletions which are among the evolutionary hotspot of the organelle are believed to be use as a tool to solve taxonomic issues among taxa that their phylogenetic relationships are unresolved. Phylogenetic relationship generated from single or combination of few genes are being replaced by the ones constructed from the whole genome as a result of new DNA sequencing methods such as next generation sequencing (NGS). The introduction of next generation sequencing has increased the availability of data for solving phylogenetic relationship issues. However, in spite of its importance, the approach is not fully and well utilize by researchers in plant systematic studies [27–29]. One of the most important benefits of next generation sequencing technique is that it generate very high amount of sequences compared with sanger sequencing technique. Additionally, the platform used in next generation sequencing like Illumina is very cheap process [30]. This approach has been used to generate huge number of data for inferring phylogenetic relationship in different taxonomic levels inference [31–34].
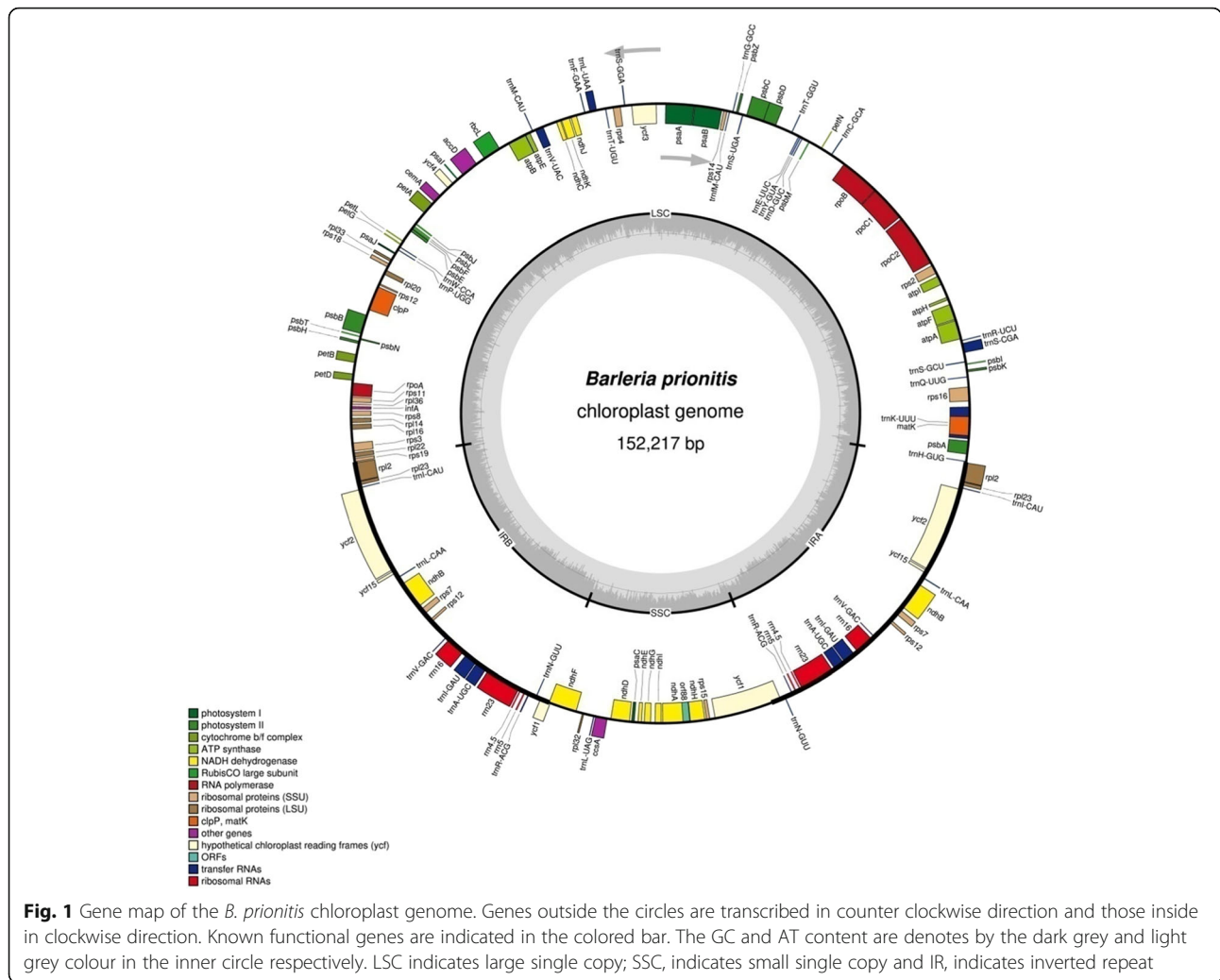
With the advent of next generation sequencing, importance of plastome sequence in resolving phylogenetic relationships and the great number of genera in Acanthceae, only plastome of few genera have been sequenced and no phylogenomic studies have been conducted for the family.

In this research, we sequenced and characterized the plastome of *Barleria prionitis* and compared the genome with cp genomes from Acanthoideae species. We used data from the whole chloroplast genome of 8 genera belonging to the Acanthoideae to reveal their tribal positions. This is as a result of incongruent of previous studies in placing the genera in their respective tribes [35]. placed Barlerieae and Andrographideae as sub tribes under the tribe Justicieae, this classification has been reported by other student of Acanthaceae [27]. classify the sub family Acanthoideae into two tribes, placing Ruelliinae, Justiciinae, Andrograpiinae and Barleriinae under the tribe Ruellieae. Findings of recent studies by McDade and her colleagues using molecular data contradict with previous classifications. Therefore, there is need to use complete chloroplast genome to address the correct placement of the genera into their respective tribes. The result of this study will be useful for developing makers, provide resources for evolutionary studies and authentication of *B. prionitis* and the inference of phylogenetic relationships within Acanthoideae.

Results

Characteristics of *B. prionitis* chloroplast genome

The complete plastome sequence of *B. prionitis* was reported to be 152,217 bp in size and has a structural organization of quadripartite containing a large single copy (LSC, 83, 772 bp), a pair of inverted repeat (IRa and IRb 25, 321 bp each) and small single copy (SSC, 17, 803 bp) (Fig. 1 and Table 1). Composition of Adenine-Thymine and Guanine-Cytosine content in *B. prionitis* was 61.7 and 38.3%, respective whereas the IRA, IRB, SSC and LSC regions have, 67.4 and 32.6%, 56.5 and 43.5%, 56.4 and 43.6%, and 63.6% and 36.4, respectively (Table 1). The inverted repeat region have higher GC

Fig. 1 Gene map of the *B. prionitis* chloroplast genome. Genes outside the circles are transcribed in counter clockwise direction and those inside in clockwise direction. Known functional genes are indicated in the colored bar. The GC and AT content are denotes by the dark grey and light grey colour in the inner circle respectively. LSC indicates large single copy; SSC, indicates small single copy and IR, indicates inverted repeat

content of 49% compared with the SSC and LSC regions with 32.6 and 36.4% respectively (Table 1). In terms of the size of the coding and non coding region, the protein coding regions is 79, 950 pb in length whereas the non coding which includes the intergenic spacer and introns have 72, 267 bp.

Table 1 Nucleotide composition in the complete plastome sequence of *B. prionitis*

Region		T(U) (%)	C (%)	A (%)	G (%)	Length (bp)
cp Genome		31.2	19.5	30.5	18.8	152,217
LSC		32.4	18.7	31.2	17.7	83,772
SSC		33.6	17.1	33.8	15.5	17,803
IRA		28.2	22.5	28.2	21.0	25,321
IRB		28.2	21.0	28.3	22.5	25,321
	1st Position	30	20.4	30.4	19.0	50,739
	2nd Position	32	18.7	31.3	18.0	50,739
	3rd Position	31	19.5	29.8	19.4	50,739

The complete chloroplast genome of *B. prionitis* contained 113 different genes out of which 18 are duplicated in the IRA and IRB region, totaling 131 genes. The number of rRNA genes, tRNA genes and protein-coding genes in the genome are 4, 30 and 80, respectively (Fig. 1 and Table 2). Four rRNA, seven protein coding and tRNA genes are located in the pair of the inverted repeat region of the plastome whereas the large single copy region harbored 62 protein-coding sequence and 22 tRNA genes, the remaining one tRNA and 12 protein coding genes are located in the single copy region. Among the genes coding for protein, many of them started with the codon ATG while few starts with other codon such as ACG and GTG, this is also reported in other chloroplast genome of angiosperms.

The chloroplast genome of *B. prionitis* is found to have intron in some of the genes, like in other species in the Lamiales order [36, 37]. Out of the 113 different genes, 14 of them contain intron (Table 3), six tRNAs and eight protein-coding genes. Four of the genes with

Table 2 Genes present in the chloroplast genome of *B. prionitis*

Category	Group of genes	Name of genes
RNA genes	ribosomal RNA genes (rRNA)	rrn5, rrn4.5, rrn16, rrn23
	Transfer RNA genes (tRNA)	trnH-GUG, trnK-UUU[a], trnQ-UUG, trnS-GCU, trnS-CGA[a], trnR-UCU,trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnfM-CAU, trnG-GCC, trnS-GGA, trnL-UAA[a], trnT-UGU, trnF-GAA, trnV-UAC[a], trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU[c], trnL-CAA[c], trnV-GAC[c], trnI-GAU[a,c], trnA-UGC[a,c], trnR-ACG[c], trnN-GUU[c], trnL-UAG,
Ribosomal proteins	Small subunit of ribosome	rps2, rps3, rps4, rps7[c], rps8, rps11, rps12[c], rps14, rps15, rps,16[a], rps18, rps19
Transcription	Large subunit of ribosome	rpl2[a,c], rpl14, rpl16, rpl20, rpl22, rpl23[a], rpl32, rpl33, rpl36
	DNA dependent RNA polymerase	rpoA, rpoB, rpoC1[a], rpoC2
Protein genes	Photosystem I	psaA, psaB, psaC, psaI,psaJ,ycf3[b]
	Photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	Subunit of cytochrome	petA, petB, petD, petG, petL, petN
	Subunit of synthase	atpA, atpB, atpE, atpF[a], atpH, atpI
	Large subunit of rubisco	rbcL
	NADH dehydrogenase	ndhA[a], ndhB[a,c], ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
	ATP dependent protease subunit P	clpP[b]
	Chloroplast envelope membrabe protein	cemA
Other genes	Maturase	matK
	Subunit acetyl-coA carboxylase	accD
	C-type cytochrome systhesis	ccsA
	Hypothetical proteins	ycf2[c],ycf4, ycf1[c]
	Component of TIC complex	ycf[c]

[a] Gene with one intron, [b] Gene with two intron and [c] Gene with copies

Table 3 Genes with intron in the *B. prionitis* chloroplast genome and length of exons and introns

| Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
rps16	LSC	37	865	228		
atp F	LSC	143	664	470		
rpoC1	LSC	431	786	1619		
ycf3	LSC	128	697	227	750	152
clpP	LSC	68	747	290	640	227
rpl2	IR	392	676	434		
ndhB	IR	776	680	755		
ndhA	SSC	551	1082	539		
trnK-UU	LSC	36	2460	37		
trnS-CGA	LSC	31	667	59		
trnL-UAA	LSC	36	487	49		
trnV-UAC	LSC	37	595	36		
trnI-GAU	IR	41	938	34		
trnA-UGC	IR	37	806	34		

intron viz.: *ndhB, trnA-UGC, trnI-GAU* and *rpl2* are situated in the inverted repeat region and the other 12 in the large single copy region. *clpP* and *ycf3* are the only genes with two intron, while the other 12 genes have one intron, this is consistent with that of *S. cusia* [36]. *trnK-UUU* is the gene with longest intron with 2460 bp because of the situation of *matK* in the gene.

The frequency of the codon usage present in the plastome of *B. prionitis* was computed using the nucleotide sequence of protein-coding genes and tRNA genes 100,319 bp, the result is presented in Table 4, the results showed the genes in the plastome are encoded by 33, 436 codons. The codons that codes

for the amino acids Leucine appears more frequently in the genome 3286 (9.83%) (Fig. 2), comparable to that of *Ailanthus altisssima* and the ones coding for Trp have the lowest 622 (1.86%) in the plastid sequence. Guanine-Cytosine ending are more common than the Adenine-Thymine ending, this is incongruent with other cp genome sequence [38–40]. The result of the analysis show that there is low codon usage bias in the plastome sequence of *B.prionitis* (Table 4). 29 codons have RSCU values greater than 1 and all of them are characterized with Adenine-Thymine ending while for 30 codons, were less than 1 and are all of Guanine-Cytosine ending. The amino acids

Table 4 Codon – anticodon recognition patterns and codon usage of the *J. flava* chloroplast genome

Codon	Amino Acid	Count	RSCU	tRNA	Codon	Amino Acid	Count	RSCU	tRNA
UUU	Phe	1278	1.18	trnF-GAA	UAU	Tyr	964	1.43	trnY-GUA
UUC	Phe	882	0.82		UAC	Tyr	384	0.57	
UUA	Leu	704	1.29	trnL-UAA	UAA	Stop	556	1.02	
UUG	Leu	717	1.31	trnL-CAA	UAG	Stop	484	0.89	
CUU	Leu	660	1.21	trnL-UAG	CAU	His	492	1.29	trnH-GUG
CUC	Leu	423	0.77		CAC	His	268	0.71	
CUA	Leu	477	0.87		CAA	Gln	685	1.38	trnQ-UUG
CUG	Leu	302	0.55		CAG	Gln	309	0.62	
AUU	Ile	1149	1.26	trnI-GAU	AAU	Asn	1046	1.39	trnG-GUU
AUC	Ile	788	0.86		AAC	Asn	463	0.61	
AUA	Ile	801	0.88	trnI-CAU	AAA	Lys	1253	1.29	trnK-UUU
AUG	Met	706	1	trnM-CAU	AAG	Lys	686	0.71	
GUU	Val	606	1.5	trnV-GAC	GAU	Asp	721	1.45	trnD-GUC
GUC	Val	258	0.64		GAC	Asp	273	0.55	
GUG	Val	292	0.72		GAA	Glu	943	1.38	trnE-UUC
GUA	Val	462	1.14	trnV-UAC	GAG	Glu	420	0.62	
UCU	Ser	704	1.45	trnS-GGA	UGU	Cys	443	1.19	trnC-GCA
UCC	Ser	447	0.92		UGC	Cys	301	0.81	
UCG	Ser	389	0.8		UGA	Stop	595	1.09	
UCA	Ser	614	1.26	trnS-UGA	UGG	Trp	622	1	trnW-CCA
CCU	Pro	416	1.24	trnP-UGG	CGU	Arg	266	0.69	trnR-ACG
CCC	Pro	283	0.85		CGC	Arg	148	0.39	trnR-UCU
CCA	Pro	393	1.17		CGA	Arg	436	1.14	
CCG	Pro	247	0.74		CGG	Arg	294	0.77	
ACU	Thr	428	1.18		AGA	Arg	761	1.98	
ACC	Thr	311	0.86		AGG	Arg	399	1.04	
ACG	Thr	242	0.67	trnT-GGU	AGU	Ser	457	0.94	trnS-GCU
ACA	Thr	471	1.3	trnT-UGU	AGC	Ser	302	0.62	
GCU	Ala	349	1.34	trnA-UGC	GGU	Gly	520	1.05	trnG-GCC
GCC	Ala	206	0.79		GGC	Gly	290	0.59	
GCA	Ala	301	1.16		GGA	Gly	670	1.36	
GCG	Ala	186	0.71		GGG	Gly	493	1	trnG-UCC
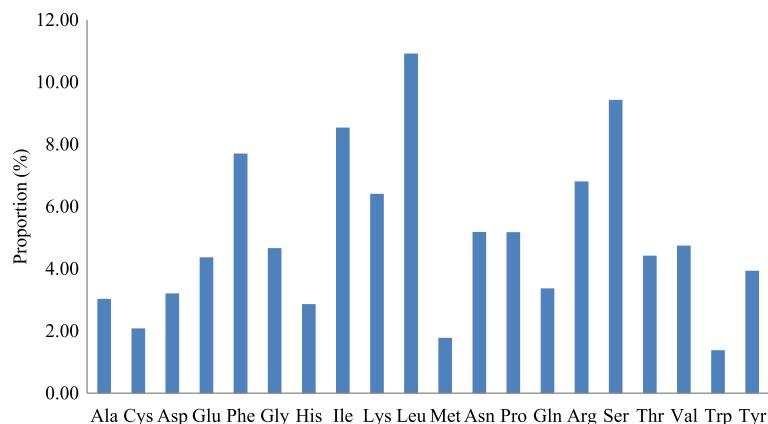
Fig. 2 Amino acids frequencies in *B. prionitis* chloroplast genome protein coding sequences

Tryptophan and Methionine have RSCU value of 1 hence they don't have codon bias.

The prediction of RNA editing sites present in the plastome sequence of *B. priniotis* was done by means of PREP suite. The first codon of the first nucleotide was used in all the analysis. The results as shown in (Table 5) showed that most of the conversions in the codon positions are from Serine to Leucine. Generally, the editing sites observed in the plastome were 61 which are distributed between the 19 protein-coding genes. *psaB* is found to have the highest number of editing site (13 sites) followed by *ndhB* (9 sites), *rpoB* (6 site) and *rpl20, accD, rps, atpI, rpl2, rpoA* have the lowest number of editing site with 1 editing site each. Nine (9) RNA editing site in *ndhB* has been confirmed in the plastome of other species [41–43]. Conversions of proline to serine were observed, which involves the changing of the amino acids in the RNA editing site from apolar to polar group. Genes such as *petD, ndhC, atpB, clpP, ndhE, petL, ndhG, petG* and *ccsA* among others do not possess RNA editing site in their first codon of the nucleotide.

Long repeats

Repeat sequence in the chloroplast genome of *B. prionitis* were screen using REPuter programme with default settings, the programme revealed that only three types of repeats were present in the genome viz. Palindromic, forward and reverse, the complement repeat is not detected within the plastome (Table 6). The result revealed 18 palindromic repeats, 25 forward repeats and 6 reverse repeats (Table 6). Most of the repeats size are between 20 and 29 bp (78.6%), followed by 10–19 bp (10.20%) whereas 40-49 bp are the least (4.08%). In all, there are 49 number repeats in *B. priniotis* plastome. In the first location, 65.30% of the repeats are contained in the non coding region; this is comparable to the cp genome of *Fagopyrum dibotrys* [44]. Eight repeats were located in the tRNA (16.32%), the other 9 repeats (18.36%) are

situated in the protein coding genes in particular *rpl2, ndhA, ycf1, ndhC,* and *ycf2*. Among the protein coding genes *ycf2* contained 2 forward palindromic and repeats.

The rate of repeats among eight Acanthoideae plastomes was compared, the results indicates that complement, palindromic, reverse and forward type of repeats occurred in the plastome of *J. flava, A. paniculata, S. cusia, B. ciliaris* and *R. breedlovei*, whereas no complement repeats detected in the cp genomes of *B. prionitis, E. attenuatus* and *A. knappiae* (Fig. 3). *S. cusia, B. ciliaris* and *A. paniculata* are found to have high frequency of palindromic repeats (23) and *J. flava* is found to have the least (16). *R. breedlovei, S. cusia* and *A. paniculata* have15 forward repeats in their plastome and the frequency of reverse repeats is identical in the plastome of *A. paniculata, S. cusia* and *J. flava*. Complement repeat is absent in *B. prionitis, E. attenuates, A. knappiae* and is the least repeat in the plastome of *J. flava, A. paniculata, B. ciliaris, R. breedlovei* and *S. cusia*.

Microsatellite analysis

Microsatellites (SSRs) are short repeat of nucleotide sequences (1-6 bp) that are distributed throughout genome. This short repeats are used as important makers for evolutionary studies of plants [45]. In this research, a total number of 98 microsatellites were identified in the chloroplast genome of *B. priniotis* (Table 7). Most of the microsatellites in the plastome are mononucleotide (83.67%) and majority of them are polythymine 58.53% followed by poly A (polyadenine) 40.24%, only one Poly G (polyguanine (1.21%) is present where as no poly C detected in the genome. Among dinucleotide only 5 repeats were detected, TA repeated four times and AT only once. Considering sequence complimentary, two trinucleotide AAG/CTT and AAT/ATT, four tetra AAAC/GTTT, AAAG/CTTT, AAAT/ATTT, AATC/ATTG and only one penta AAATGG/ATTTCC were detected in the genome (Fig. 4a) whereas no

Table 5 Predicted RNA editing site in the *B. prionitis* chloroplast genome

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Amino Acid Conversion	Score
accD	722	241	TCG = > TTG	S = > L	0.8
atpF	791	264	CCC = > CTC	P = > L	1
	914	305	TCA = > TTA	S = > L	1
atpI	620	207	TCA = > TTA	S = > L	1
matK	469	157	CAC = > TAC	H = > Y	1
	661	221	CAT = > TAT	H = > Y	1
	1264	422	CAT = > TAT	H = > Y	1
ndhA	341	114	TCA = > TTA	S = > L	1
	566	189	TCA = > TTA	S = > L	1
Gene	**Nucleotide Position**	**Amino Acid Position**	**Codon Conversion**	**Amino Acid Conversion**	**Score**
	1073	358	TCC = > TTC	S = > F	1
ndhB	149	50	TCA = > TTA	S = > L	1
	467	156	CCA = > CTA	P = > L	1
	586	196	CAT = > TAT	H = > Y	1
	737	246	CCA = > CTA	P = > L	1
	746	249	TCT = > TTT	F = > F	1
	830	277	TCA = > TTA	S = > L	1
	836	279	TCA = > TTA	S = > L	1
	1292	431	TCC = > TTC	S = > F	1
	1481	494	CCA = > CTA	P = > L	1
ndhD	2	1	ACG = > ATG	T = > M	1
	878	293	TCA = > TTA	S = > L	1
ndhF	124	42	CTT = > TTT	L = > F	1
	290	97	TCA = > TTA	S = > L	1
	1504	502	CTT = > TTT	L = > F	1
petB	424	142	CGG = > TGG	R = > W	1
	617	206	CCA = > CTA	P = > L	1
psaB	88	30	CTT = > TTT	L = > F	1
	193	65	CTT = > TTT	L = > F	1
	422	141	TCT = > TTT	S = > F	1
	430	144	CCT = > TTT	P = > F	0.86
	431	144	CCT = > TTT	P = > F	0.86
	544	182	CTT = > TTT	L = > F	1
	1090	364	CTT = > TTT	L = > F	1
	1277	426	CCT = > CTT	P = > L	1
	1279	427	CTT = > TTT	L = > F	0.86
	1546	516	CTT = > TTT	L = > F	1
	1961	654	TCT = > TTT	S = > F	1
	1993	665	CTC = > TTC	L = > F	1
	2096	699	CCT = > CTT	P = > L	1
psbE	110	37	TCG = > TTG	S = > L	1
	118	40	CCG = > TCG	P = > S	1
	146	49	GCC = > GTC	A = > V	1
	148	50	CTC = > TTC	L = > F	1

Alzahrani *et al. BMC Genomics*   (2020) 21:393

Page 8 of 19

Table 5 Predicted RNA editing site in the *B. prionitis* chloroplast genome *(Continued)*

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Amino Acid Conversion	Score
rpl2	596	199	GCG => GTG	A => V	0.86
rpl20	308	103	TCA => TTA	S => L	0.86
rpoA	887	296	TCG => TTG	S => L	1
rpoB	473	158	TCA => TTA	S => L	0.86
	551	184	TCA => TTA	S => L	1
	566	189	TCG => TTG	S => L	1
	593	198	GCT => GTT	A => V	0.86
	1289	430	ACC => ATC	T => I	0.86
	2426	809	TCA => TTA	S => L	0.86
rpoC2	2287	763	CGG => TGG	R => W	1
	3121	1041	CGC => TGC	R => C	0.8
	3725	1242	TCA => TTA	S => L	0.86
rps2	248	83	TCA => TTA	S => L	1
rps8	113	38	ACT => ATT	T => I	1
	119	40	CCG => CTG	P => L	1
	257	86	ACC => ATC	T => I	0.86
rps14	80	27	TCA => TTA	S => L	1
	149	50	TCA => TTA	S => L	1

hexanucleotide repeat detected. The majority of the microsatellites are found in the intergenic spacer region (Fig. 4b) (62.24%) and the coding region contained the least (33.67%). The majority of the repeats were located in the large single copy region (70.40%) and the single copy region contained the lowest frequency of repeat (9.18%) in the plastome.

The rate of occurrence of SSRs among the plastomes of the eight members of Acanthoideae was compared (Fig. 4c); the comparison indicate high frequency of mononucleotides across all the plastomes. *E. attenuatus* and *A. paniculata* had the highest number of mononucleotide with 107 and 104 respectively. Pentanucleotides were not found in the plastome of *B. prionitis, E. attenuatus, A. knappiae, B. ciliaris* and *R. breedlovei* while hexanucleotide were only present in *B. prionitis, R. breedlovei* and *A. knappiae*.

Comparative analysis of Justicia flava chloroplast to other Acanthaceae genomes

The plastome sequences of eight Acanthaceae species namely (*B. prionitis, J. flava, B. ciliaris, A. paniculata, E. attenuatus, R. breedlovei, A. knappiae* and *S. cusia* were compared. To check the level of nucleotide sequence variation between the sampled plastomes of Acanthoideae species, the programe mVISTA was used to aligned the sequences with the annotation of *B. prionitis* as reference. Result of the alignment indicates that the plastomes are extremely conserved, however some level of variations were detected. The pair of the iverted repeat is highly conserved than the small single copy region and large single copy region. Additionally, the protein-coding genes are highly conserved than the non coding region, mostly the integernenic spacer regions. The intergenic spacer regions with high level of variation within the gemone are *trnL – trnA, trnH-GUG – psbA, trnC – petN, trnL – trnF, accD – psaI, rps12- trnV, rps15 – ycf1, rps16 – trnQ* (Fig. 5). The protein coding genes that showed sequence divergence are *ycf2, psbL, atpE, rbcL, petB, petA,* and *atpF*.

The plastome sequence of flowering plant is reported to have generally been conserved [46], although there is a little variations in size and boundries of the single copy and inverted repeats as a results of the evolutionary happenings such as contraction and expansion in the plastome architecture [47, 48]. The comparison between the invterted repats and single copy regions boundries in the eight plastome of Acanthaceae (*B. prionitis, B. ciliaris, A. paniculata, E. attenuatus, R. breedlovei, J. flava, S.cusia* and *A. knappiae* are presented in (Fig. 6). There is a little variation in the boundaries of the IR-SSC and IR-LSC of the plastomes (Fig. 6),the *rps19* is located in LSC region of *B. prionitis, B. ciliaris, A. paniculata* and *A. knappiae*. The following genes *trnH, rps19, ycf1* and *ndhF* are located at the junction of IR-SSC and IR-LSC of *J. flava* and *E. attenuatus* plastomes slightly variation in number of nucleotides (Fig. 6). In the SSC/IRb border of the eight plastomes, *ycf1* and *ndhF* genes are found.

Table 6 Repeat sequences present in the *B. prionitis* chloroplast genome

S/No	Repeat size	Repeat type	Repeat Position 1	Repeat Location 1	Repeat Position 2	Repeat Location 2	E-value
1	41	F	97,869	*rpl2*	118,701	*ndhA*	1.35E-15
2	41	P	118,701	*ndhA*	138,043	*rps12*	1.35E-15
3	34	F	45,870	IGS	45,903	IGS	2.21E-11
4	32	F	42,498	*ycf3* Intron	42,528	*ycf3* Intron	3.53E-10
5	30	P	7813	IGS-*trnS-GCU*	44,688	*trnS-GGA*	5.65E-09
6	29	F	345	IGS	71	IGS	2.26E-08
7	27	F	45,851	IGS	45,867	IGS	3.62E-07
8	26	P	87,326	*ycf2*	87,326	*ycf2*	1.45E-06
9	26	F	87,326	*ycf2*	148,601	*ycf2*	1.45E-06
10	26	P	111,582	IGS	111,582	IGS	1.45E-06
11	26	F	122,197	IGS	122,222	IGS	1.45E-06
12	26	P	148,601	*ycf2*	148,601	*ycf2*	1.45E-06
13	24	F	45,854	IGS	45,903	IGS	2.32E-05
14	23	F	43,353	IGS	97,871	IGS	9.26E-05
15	23	F	43,353	IGS	118,703	IGS	9.26E-05
16	23	P	43,353	IGS	138,059	IGS	9.26E-05
17	23	F	59,543	IGS	59,564	IGS	9.26E-05
18	23	R	65,867	IGS	65,867	IGS	9.26E-05
19	23	F	70,995	IGS	71,017	IGS	9.26E-05
20	22	F	9299	*trnG-GCC*	36,153	*trnG-UCC*	3.70E-04
21	22	P	30,692	IGS	30,692	IGS	3.70E-04
22	22	F	90,910	IGS	90,928	IGS	3.70E-04
23	22	P	90,910	IGS	145,003	*ycf2*	3.70E-04
24	22	P	90,928	IGS	145,021	*ycf2*	3.70E-04
S/No	Repeat size	Repeat type	Repeat Position 1	Repeat Location 1	Repeat Position 2	Repeat Location 2	E-value
25	22	P	93,326	IGS	93,352	IGS	3.70E-04
26	22	F	93,326	IGS	142,579	IGS	3.70E-04
27	22	F	93,352	IGS	142,605	IGS	3.70E-04
28	22	P	142,579	IGS	142,605	IGS	3.70E-04
29	22	F	145,003	*ycf2*	145,021	*ycf2*	3.70E-04
30	21	F	82	IGS	103	IGS	1.48E-03
31	21	F	7819	*trnS-GCU*	35,296	*trnS-UGA*	1.48E-03
32	21	P	35,296	*trnS-UGA*	44,691	*trnS-GGA*	1.48E-03
33	21	F	36,373	*trnfM-CAU*	66,402	*trnP-UGG*	1.48E-03
34	21	R	119,320	IGS	119,320	IGS	1.48E-03
35	20	F	3731	IGS	111,632	IGS	5.93E-03
36	20	R	30,031	IGS	30,031	IGS	5.93E-03
37	20	R	45,796	IGS	45,796	IGS	5.93E-03
38	20	R	49,989	*ndhC*	49,989	*ndhC*	5.93E-03
39	20	P	51,944	*trnV-UAC*	102,610	*trnA-UGC*	5.93E-03
40	20	F	51,944	*trnV-UAC*	133,323	*trnA-UGC*	5.93E-03
41	20	P	56,548	IGS	56,548	IGS	5.93E-03
42	20	F	57,115	IGS	57,135	IGS	5.93E-03
43	20	P	73,944	IGS	73,967	IGS	5.93E-03

Table 6 Repeat sequences present in the *B. prionitis* chloroplast genome *(Continued)*

44	20	P	123,426	*ycf1*	123,426	*ycf1*	5.93E-03
45	19	P	70,448	IGS	123,029	*ycf1*	2.37E-02
46	18	F	249	IGS	272	IGS	9.48E-02
47	18	P	1946	IGS	42,695	*ycf3* Intron	9.48E-02
48	18	R	7245	IGS	7245	IGS	9.48E-02
49	18	F	7884	*trnS-GCU*	35,366	*trnS-UGA*	9.48E-02

Positioning of *ycf2* gene in the IRb/LCS border is observed only in the genome of *R. breedlovei* where as *E. attenuatus* plastome also has distinctive structural variation of having *rpl22* in junction of IRb/LSC. The gene *ndhF* was found to have 36 bp, 109 bp, 40 bp and 41 in the IRb region in *B. prionitis, E. attenuatus, A. paniculata* and *A. knappiae* respectively where as *trnH* of *E. attenuatus* and *J. flava* is positioned at IRa/LSC border .

Divergence of protein coding gene sequence

The dN/dS ratio and rates of nonsynonymous (dN) substitution and synonymous (dS) were calculated using DNAsp among the plastome of eight species of Acanthoideae to detect the protein-coding genes that were under selective pressure. The results revealed that the dN/dS ratio is < 1 in most of the genes with the exception of *atpF, petD, psbZ* and *rpl20 of B. prionitis* vs *E. attenuatus, petB, petD, rpl16, rpoC* and *rps16 of B. prionitis* vs *A. paniculata, petD, psbZ, rpl16, rps7of B. prionitis* vs *A. knappieae, psbZ of B. ciliaris, rpl32*and *ycf3* of *B. prionitis* vs *J. flava* having 1.16, 2.08, 2.76 and 1.72, 2.74, 2.30, 2.71, 1.65, 1.30 and 1.61, 2.70, 2.41, 2.76 and 1.61, 1.19, 1.45 and 1.32 respectively (Fig. 7). This shows that the majority of the genes undergo negative selection only few of them were under positive selection. The values of synonymous (dS) rate ranged from 0.01 to 0.38 in all the genes (Fig. 7). Some of the genes including *psaJ, atpH, ndhC, psaI, psbE, rpl2, psbH, psbI, psbL* and *psbF* showed no nonsynonymous changes.

Phylogenetic analysis

To determine the phylogenetic relationship and tribal positions of the nine species of Acanthaceae, we used the plastome of the eight species to reconstruct phylogenetic tree. The phylogenetic analyses were performed using Maximum likelihood and Bayesian inference (BI) with *Erythranthe lutea, Scrophularia dentate, Lysionatus pauciflorus* and *Tanaecium tetragonolobum* as outgroup. The resulting tree from Bayesian inference (BI) and Maximum likelihood analyses were congruent with high support PP, 1.0 and MP, 100 in all relationships (Fig. 8). All the nine species clustered in one clade with strong support and are divided into two major sub clades. Sub clade 1 which is monophyletic includes *A. knappiae* and *B. ciliaris* (Acantheae) is sister to large clade 2 containing Ruellieae, Barlerieae, Justicieae, Andrographideae. Within the second clade Justicieae and Ruellieae are sister taxa as well as Barlerieae and Andrographideae.

Discussion

In this study, we sequenced the plastome sequence of *B. prionitis* using Illumina sequencing technology. This is a new approach of obtaining cp genome without prior isolation of the cpDNA and it has been used in several studies. The analysis of the cp genome revealed that the genome has a quadripartite structure; with a pair of inverted repeats regions (IRa and IRb) separated by small single copy region (SSC) and large single copy region (LSC). The organization and structure of the *B.*
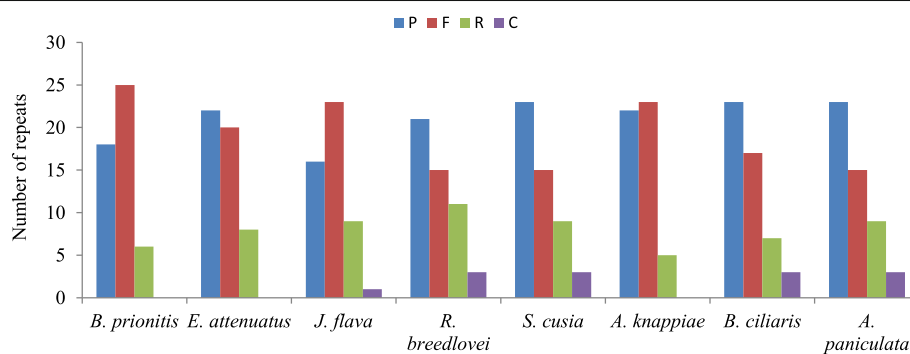
Fig. 3 Number of different repeats in four chloroplast genome of Acanthaceae. P = palindromic, F = forward, R = reverse and C = complement
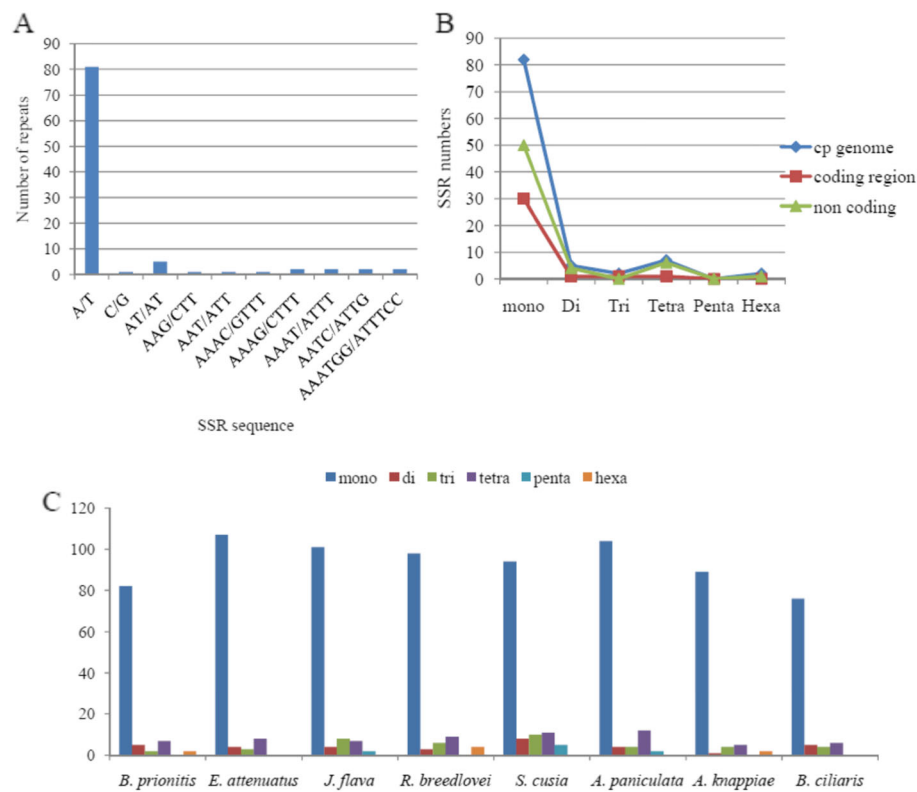
Table 7 Simple sequence repeats in the chloroplast genome of *B. prionitis*

cpSSR ID	Repeat Motif	Length (bp)	No. of Repeats	SSR start position
1	(A)8	8	22	4, 148; 11, 237; 15,924; 18,270 (*rpoC2*); 21,932 (*rpoC1*); 22,378; 44,464; 46,615; 59,100; 62,380 (*petA*); 64,357 (*psbF*); 68,879; 70,310; 70,923; 74,985; 96,279; 110,695 (*ndhF*); 113,108 (*ccsA*); 115,491 (*ndhD*); 123,247 (*ycf1*); 124,939 (*ycf1*); 126,594 (*ycf1*)
2	(A)9	9	8	3363 (*matK*); 89, 137 (*ycf2*); 28,351; 30,176; 68,148; 89,137 (*ycf2*); 112,433; 133,945
3	(A)10	10	3	6496; 70,448; 115,641
4	(G)8	8	1	58, 275 (*accD*)
5	(T)8	8	21	9447; 9599; 9655; 30,954; 35,949; 46,646; 46,783; 59,282; 59,842 (*ycf4*); 68,159; 70,217; 70,286; 73,675; 76,792 (pet D); 83,687 (rps19); 109,738 (*ndhF*); 112,390; 113,446 (*ccsA*); 114,650 (*ndhD*); 124,356 (*ycf1*); 139,668
6	(T)9	9	16	7179 (*psbK*); 12,618; 13,606; 16,227 (*rpoC2*); 35,155; 45,651; 54,687; 65,875; 67,203; 83,719 (*rps19*); 102,001; 122,650 (*ycf1*); 123,040 (*ycf1*); 124,647 (*ycf1*); 124,977 (*ycf1*); 146,809 (*ycf2*)
7	(T)10	10	3	42,062; 54,268 (*atpB*); 77,754 (*rpoA*)
8	(T)11	11	4	18,127 (*rpoC2*); 29,006; 43,047; 50,631
9	(T)12	12	2	7567; 11,941;
10	(T)13	13	1	81,706
11	(T(16)	16	1	30,034
12	(AT)5	5	1	19,500 (*rpoC2*)
13	(TA)5	5	4	45,796; 45,807; 45,823; 58,964
14	(ATA)4	4	1	54,322
15	(TTC)4	4	1	34,936 (*psbC*)
16	(AAAC)1	1	1	67,703
17	(AATA)1	1	1	114,259 (*ndhD*)
18	(AATC)1	1	1	121,984
19	(AGAA)1	1	1	9020
20	(ATAA)1	1	1	54,335
21	(GAAA)1	1	1	60,503
22	(GATT)1	1	1	5638
23	(AATGGA)1	1	1	99,217
24	(TTTCCA)1	1	1	136,718

prionitis cp genome is similar to other sequenced Acanthaceae cp genomes [49, 50]. Notably, there is high variation in terms of genome size and organization between *B.prionitis* and *S. cusia*, this is as a result of IR contraction. The size of the genome 152,217 bp is comparable to other sequenced cp genome of Acanthaceae species, longer than *A. paniculata* [51], *R. breedlovei* [50] and *S. cusia* [52] shorter than *E. attenuatus* [49]. The size of the genome in all the studied species is relevant to variation in the LSC region. The cp genome of *B. prionitis* was found to posses 38.3% GC content, as in *S.cusia* [52]. Additionally, *rps12* was recognized as trans-spliced gene, this was reported in other species [52–54]. The arrangement and gene contents of the *B. prionitis* cp genome is similar to other sequence cp genome of Acanthaceae [50, 51] but is different with that of *S. cusia* which has *trnH-GUG* in the inverted repeat regions and *ycf2* in the large single copy [52]. Some of the genes in the cp genome of *B. prionitis* start with ACG, GTG and

ATC codon, this phenomenon have been reported in angiosperm chloroplast genome [36, 37, 55].

Repeat elements present in cp genome are correlated with the genome recombination and rearrangements [56, 57]. The cp genome of *B. prionits* is found have low number of repeats compared to sequenced Acanthaceae plastome [47, 51, 52]. Acanthaceae plastomes contained low repeats compared with other angiosperm cp genome. Most of the repeats were located in the non coding region and *ycf* genes (*ycf1* and *ycf2*), this has been commonly observed in plastome of angiosperms [58]. Chloroplast microsatellites (cpSSRs) are short repeat in chloroplast genome inherited from a single parent, hence are often used as molecular makers in evolutionary studies such as genetic diversity, they also play role in identification of species [59–61]. cp microsatellites analysis, reveal total number of 98 SSRs in the cp genome of *B. prionits* of which most are mononucleotides, A and T. Poly A and T are reported to be the most

Fig. 4 Simple sequence repeats (SSRs) types, distribution and presence in *B. prionitis* and other representatives species from Acanthoideae (**a**) Frequency of different SSR motifs in different repeat types in *B. prionitis* chloroplast genome. **b** Number of SSR types in complete genome, protein coding regions and Non coding genes. **c** Number of different SSR type in the eight chloroplast genome of Acanthoideae

abundant repeat in cp genome of plants [62–64]. Most of the cpSSRs are located in the non coding region whereas few are located in the protein coding genes region. The microsatellite detected in this study will be useful in evolutionary studies of the genus *Barleria* as well as identification and conservation of the genus.

Variation in size among cp genome is as a result of contraction and expansion of the inverted repeats (IRs) [65]. Contraction and expansion in IRs region were observed in the cp genome *B. prionitis* and other sequenced Acanthaceae. The size of the inverted repeats ranges from 16, 328 bp in *S. cusia* to 25, 761 bp in *E. attenuatus*. Despite the similar lengths of the IR regions of *B. prionitis* and the other Acanthaceae species with the exception of *S. cusia* some level of expansion and contraction were observed. There are variation in the border of IR-SC region among the eight species compared, we identified six type of junctions based on the position of *rps19*, *rpl2* and *trnH*, which occur as a result of contraction and expansion in the inverted repeat region. Type I occurs in three species *B. prionitis*, *A. knappiae* and *A. paniculata*, one of the duplicated *rpl2* is located in the LSC region while the other is in the IRb region whereas only 1 *rps19* is present in the LSC

region. Type II was found in *E. attenuatus*, here the two *rps19* are located in the inverted repeat regions (IRa and IRb) and the *rpl22* gene is located in the LSC region. Type III pattern occurs in *S. cusia* and is characterized by having *trnH-GUG* duplicated in the inverted region. Type IV has no genes in the IRb/LSC border and was only found in *R. breedlovei*. In type V which is observed in the genome *J. flava*, some part of the *rps19* gene are located in the inverted repeat region while some are located in single copy region, another remarkable observation is that the two *rps19* are of unequal length. The last pattern, type VI occurs in *B. ciliaris* and is characterized by having *rps19* in the LSC region and *rpl2* in the IRb region. All the genomes have *ndhF* in the IRb/SSC border as well as *ycf1* in the SSC/IRa border. It is observed that there is extension of inverted repeat into the single copy region in genome of *S. cusia* which made the LSC region to have length of 93, 666 bp. Despite the conserve nature of the cp genome, some variation could be detected [65]. The positioning of *ycf1* gene in IRb, is considered a pseudogene in many flowering plant plastomes. In addition, the stop codon is absent in the *ycf1* gene sequence and this result to the differences in the distribution of genes in single copy and inverted repeat
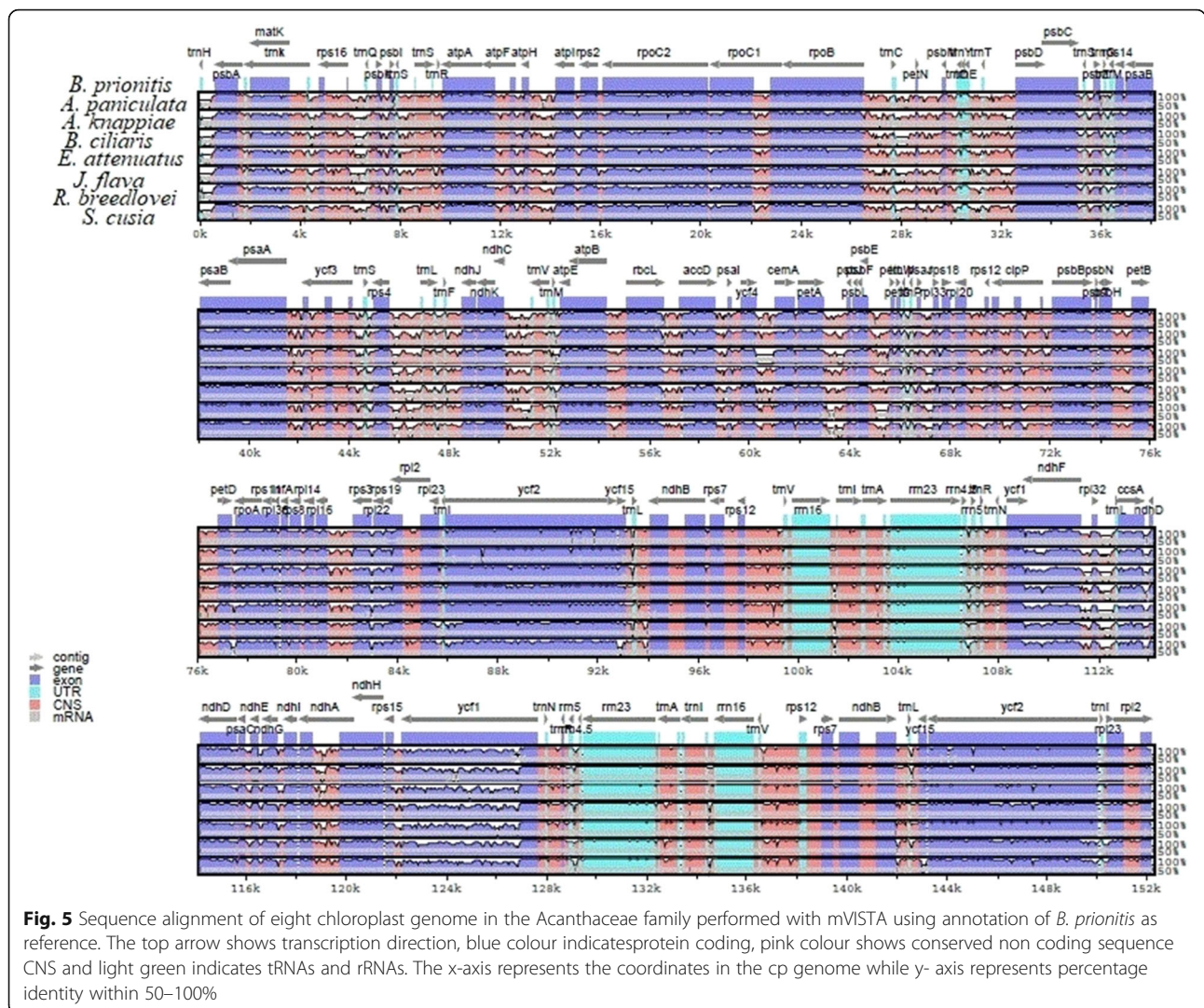
Fig. 5 Sequence alignment of eight chloroplast genome in the Acanthaceae family performed with mVISTA using annotation of *B. prionitis* as reference. The top arrow shows transcription direction, blue colour indicatesprotein coding, pink colour shows conserved non coding sequence CNS and light green indicates tRNAs and rRNAs. The x-axis represents the coordinates in the cp genome while y- axis represents percentage identity within 50–100%

borders. The result of the comparative genome analysis using mVISTA revealed that the genome is relatively conserved with some degree of variation, which mostly occurs in the non coding region as a result of insertion and deletion. The results of the alignment showed no considerable structural rearrangements, like gene relocation or inversion were detected in the plastomes. The structural rearrangement was detected in the cp genome of *S. cusia*. DNA barcodes are sequences in the genome unique to particular taxa and are used as reliable tools for identification of plants and resolving phylogenetic relationship [65, 66]. The alignment of the eight cp genome reveals variable regions which includes *trnH-GUG – psbA, rps16 – trnQ, trnC – petN, accD – psaI, clpP* intron, *trnL – trnF, rps15 – ycf1, rps12- trnV, trnL – trnA, atpE, atpF, rbcL*. These regions will be used as makers for identification of the sampled Acanthaceae species as well as resolving phylogetic relationships in the family. Most of the variable regions are located in the single

copy region particularly the large single copy, this is consistent in most angiosperms.

Synonymous (dS) and non synonymous (dN) substitution rate as well as dN/dS ration were calculated to evaluate sequence divergence and purifying selection in the protein coding genes. The result indicates low sequence divergence in most of the genes (dS < 0.1). The dN/dS analyses show that most of the protein coding genes were under negative selection, only few genes *(atpF, petD, psbZZ, rpl20, petB, rpl16, rps16, rpoC, rps7, rpl32* and *ycf3)* were under positive selection (dN/dS > 1), comparable findings were reported for other plastomes [66–68].

Complete chloroplast genome is a good resource for inferring evolutionary and phylogenetic relationships [69–71]. Many researchers have used the plastome sequence to resolve phylogenetic relationships at various taxonomic levels [72, 73]. Until this study, the phylogenetic relationships and tribal classification of Acanthaceae

Alzahrani *et al. BMC Genomics*   (2020) 21:393

Page 14 of 19
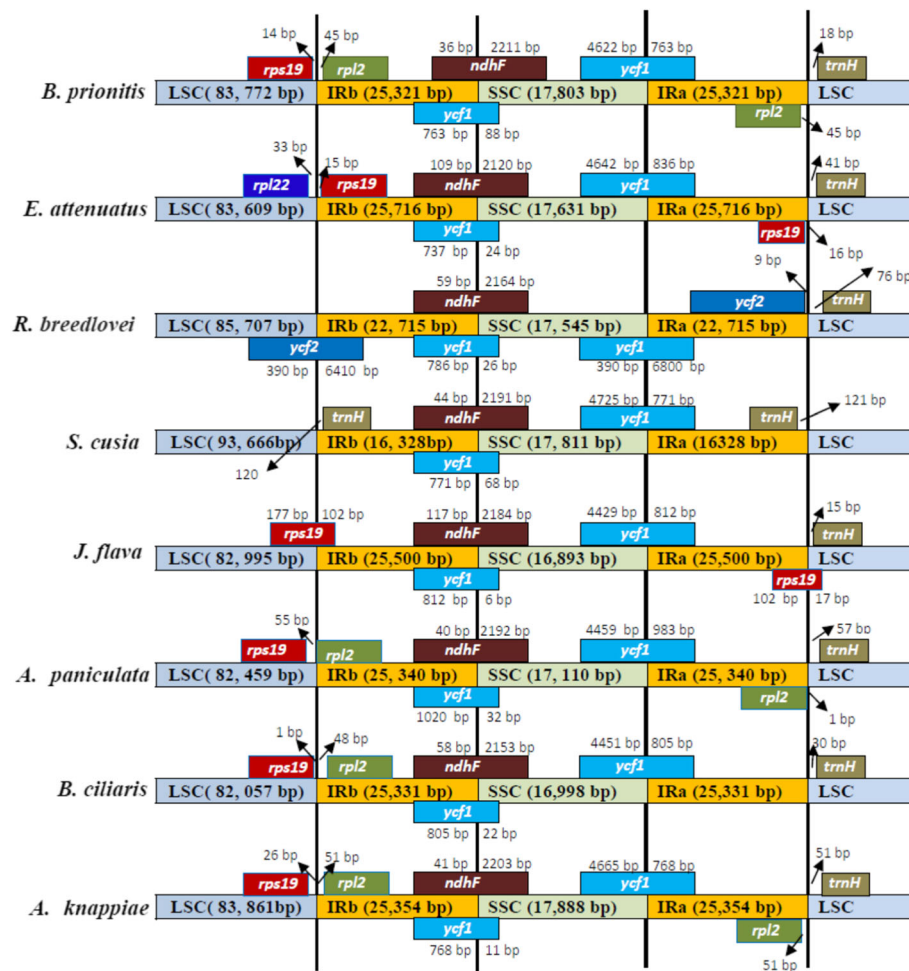
Fig. 6 Comparison of the borders of the IR, SSC and LSC regions among eight chloroplast genome of Acanthaceae
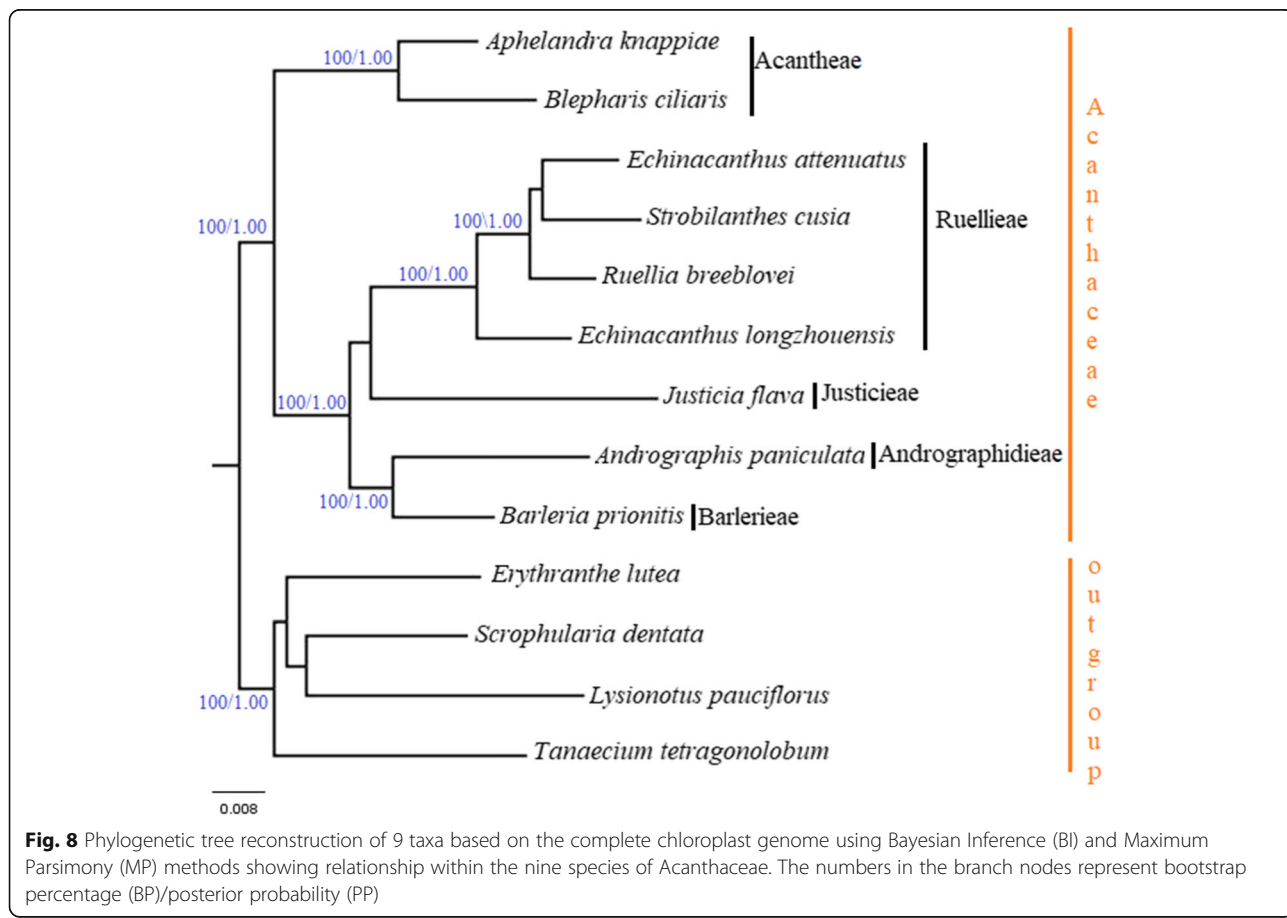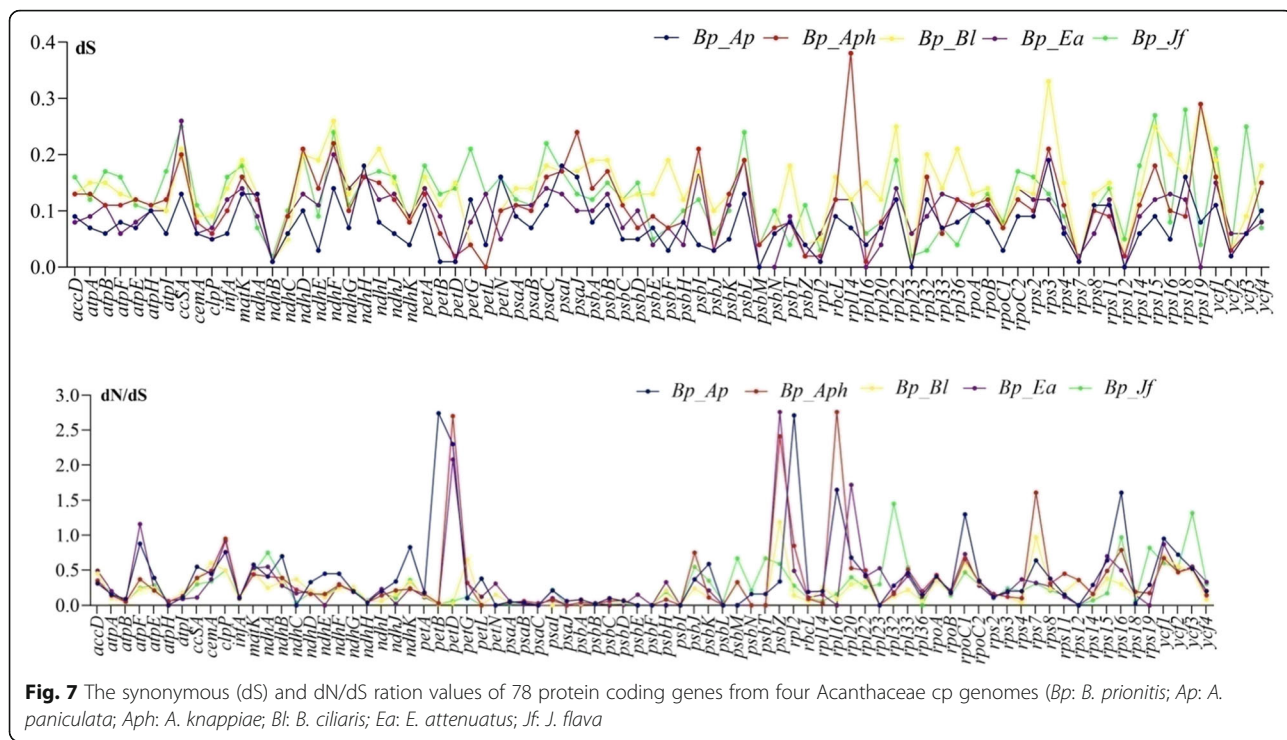
was evaluated using only few genes and the tribal classification is still required to be clarified. In this study, we used the cp genome of nine species representing the four major tribes of the Acanthoideae and reconstructed phylogenetic relationships based on maximum parsimony and Bayesian inference methods. The resulting phylogenetic tree from the two methods showed the same topology with high resolution values at the clades. The result of this study based on nine Acanthaceae taxa confirm that Acanthoideae (the retinaculate clade) are monophyletic and also confirm the sister relationship between Acantheae (non cystolith clade) and the cystolith clade, this has been reported earlier [11–13, 19] . The phylogenetic tree showed Justicieae and Ruelliae are sister taxa as reported previously [19] therefore should be regarded as separate tribes not as Justicieae or Ruelliae because the species within these two taxa are paraphyletic. The sister relationship between Andrographideae and Barlerieae is also confirm. Andrographideae and Barlerieae were placed in the tribe Justiceae as sub

tribes [35, 74]. Recently Scotland and Vollesen classified all species with cystolith under the tribe Ruellieae placing *Andrographis*, *Barleria* and *Justicia* under the sub tribes Andrographinae, Barleriinae and Justiciinae respectively. Our findings suggested that Andrographideae, Justicieae and Barlerieae should be treated as tribes not sub tribes.

Conclusion

In this study, we sequenced and reported the complete chloroplast genome of *B. prionitis*, providing valuable plastome genomic resources for the species. The plastome of *B. prionitis* has a typical gymonosperm cp genome structure and is comparable to other cp genome of Acanthaceae. Simple sequence repeats that will be used for evolutionary studies within Barleria were identified. The genome comparative analyses of 9 Acanthaceae reveal variable hotspot that could be used to develop DNA barcode for the identification of the species. These hotspots will also be useful in phylogenetic relationship

Fig. 7 The synonymous (dS) and dN/dS ration values of 78 protein coding genes from four Acanthaceae cp genomes (*Bp*: *B. prionitis*; *Ap*: *A. paniculata*; *Aph*: *A. knappiae*; *Bl*: *B. ciliaris*; *Ea*: *E. attenuatus*; *Jf*: *J. flava*

Fig. 8 Phylogenetic tree reconstruction of 9 taxa based on the complete chloroplast genome using Bayesian Inference (BI) and Maximum Parsimony (MP) methods showing relationship within the nine species of Acanthaceae. The numbers in the branch nodes represent bootstrap percentage (BP)/posterior probability (PP)

Alzahrani *et al. BMC Genomics*   (2020) 21:393

Page 16 of 19

studies of the family Acanthaceae. The study also reveals that only few genes were under positive selection. The findings of the confirmed the tribal position of major genera within Acanthoideae and suggested that Andrographideae, Justicieae and Barlerieae should be treated as tribes not sub tribes.

Methods

Plant material and DNA extraction

Plant material was collected from Makkah Taif road, Saudi Arabia (39^0 20′ 0.30″E, 21^0 45′ 33.68″N) and identified by the curator of King Abdulaziz University Herbarium, Dr. Dhafer A. Alzahrani, the voucher specimen was deposited in the herbarium of King Abdulaziz University, Jeddah, Saudi Arabia, with voucher specimen number KAU22534. Total genomic DNA was extracted from leaves using Qiagen DNA extraction Kit according to manufacturer's protocol.

Library construction, sequencing and assembly

The genomic DNA was sequenced using Illumina Hiseq 2500 platform (Novogene Technologies, Inc. Beijing, China). Raw data reads were filtered by PRINSEQ lite Ver0.20.4 [75] to get clean reads (5GB). The cp genome was assembled from the high quality clean reads using NOVOplasty2.7.2 [76] with kmer 39 using the cp genome of *Ruellia breedlovei* (KP300014.1) as reference and *ndhF* from *B. prionitis* (U12653) as seed.

Gene annotation

Dual Organellar GenoMe Annotator (DOGMA) [77] was used to annotate the genes in plastome followed by manual adjustment of the positions of start and stop codons. TrNAscan-SE2.0 [78] was used to verify tRNA genes. Organellar Genome Draw (ORGDRAW) [79] was used to circular map of plastome. The complete chloroplast genome sequence of *B. prionitis* was submitted to GenBank (Accession number MK548575).

Sequence analysis

Relative synonymous codon usage values (RSCU), base composition and codon usage were analyzed using MEGA 6.0. PREP suite [80] with cutoff value of 8.0 was used to predict the RNA editing sites in the plastome.

Repeat analysis in *B. prionitis* chloroplast genome

MIcroSAtellite (MISA) [81] was used to identify the simple sequence repeats (SSRs) with the following parameters: eight for mononucleotides, five for dinucleotides, four trinucleotides and three for tetra, penta, hexanucleotides SSR motifs. Long repeats analysis was done using the program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [80] with default parameters.

Table 8 Accession numbers of plastomes analysed in the study

Name of species	Accession Number
Aphelandra knappiae	NC_041424.1
Blepharis ciliaris	NC_046601.1
Echinacanthus attenuatus	NC_039762.1
Strobilanthes cusia	NC_037485.1
Ruellia breedlovei	KP300014.1
Justicia flava	NC_044862.1
Andrograhis paniculata	NC_022451.2
Barleria prionitis	MK548575.1
Erythranthe lutea	NC_030212.1
Scrophularia dentata	NC_036942.1
Lysionotus pauciflorus	NC_034660.1
Tanaecium tetragonolobum	NC_027955.1

Genome comparison

mVISTA [82] was used to compare the plastome using the annotation of *B. prionitis* as reference in the Shuffle-LAGAN mode [83].

Characterization of substitution rate

To detect the genes that were under selective pressure, DNAsp v5.10.01 [84] was used to analyze the synonymous (dS), nonsynonymous (dN) and dN/dS value of all the protein coding genes in sampled Acanthoideae species.

Phylogenetic analysis

For phylogenomic analysis, the cp genomes of Acanthoideae species deposited in the GenBank were recovered (Table 8). The plastome of four species of the order lamiales were also downloaded and set as out groups (Table 8). The downloaded sequences and cp genome of *B. prionitis* were aligned with MAFFT v.7 [85] and analyzed using Maximum parsimony with (PAUP version 4.0b10) [86] and Bayesian Inference with MrBayes version 3.2.6 [87].. To select the suitable model for Bayesian analysis jModelTest 3.7 [88] was used.

Author details
[1]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. [2]Department of Biology, Umaru Musa Yaradua University, Centre for Biodiversity and Conservation, Katsina, Nigeria. [3]Department of Biological Sciences, Umm Al-Qura University, Makkah, Saudi Arabia.

References

1. Bergianska.< 2015.www.angio.bergianska.se/asterids/Plantaginales/Plantagin-ales.html>. Accessed Aug 2019.
2. Olmstead RA. Synoptical Classification of the Lamiales. Version 2.4. http://depts.washington.edu/phylo/Classification.pdf. 2012.
3. APG III. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc. 2009;161:105–21.
4. Royle JF. Illustrations of the botany and other branches of the natural history of the Himalayan Mountains and of the flora of cashmere. New Delhi I: Today's and Tomorrow's Printers & Publishers; 1970. p. 296–8..
5. Solereder H. Systematic anatomy of the Dicotyledons: a handbook for Laboratories of Pure and Applied Botany, 2 v. Oxford: Clarendon Press; 1908.
6. Patil AM, Patil DA. Petiolar anatomy of some hitherto unstudies Acanthaceae. J Exp Sci. 2012;3:5–10.
7. Schonenberger J, Endress PK. Structure and development of the flowers in *Mendoncia, Pseudocalyx,* and *Thunbergia* (Acanthaceae) and their systematic implications. Int J Plant Sci. 1998;159:446–65.
8. Schonenberger J. Floral structure, development and diversity in Thunbergia (Acanthaceae*). Bot J Linn Soc. 1999;130:1–36.
9. Schwarzbach AE, McDade LA. Phylogenetic relationships of the mangrove family Avicenniaceae based on chloroplast and nuclear ribosomal DNA sequences. Syst Bot. 2002;27:84–98.
10. Wortley AH, Harris DJ, Scotland RW. On the taxonomy and phylogenetic position of Thomandersia. Syst Bot. 2007;32:415–44. https://doi.org/10.1600/036364407781179716.
11. Hedren M, Chase MW, Olmstead RG. Relationships in the Acanthaceae and related families as suggested by cladistic analysis of *rbcL* nucleotides equences. Plant Syst Evol. 1995;194:93–109.
12. Scotland RW, Sweere JA, Reeves PA, Olmstead RG. Higher- level systematics of Acanthaceae determined by chloroplast DNA sequences. Am J Bot. 1995; 82:266–75.
13. McDade LA, Moody ML. Phylogenetic relationships among Acanthaceae: evidence from non-coding *trnL-trnF* chloroplast DNA sequences. Am J Bot. 1999;86:70–80.
14. McDade LA, Masta SE, Moody ML, Waters E. Phylogenetic relationships among Acanthaceae: evidence from two genomes. Syst Bot. 2000;25:105–20.
15. Scotland RW, Endress PK, Lawrence TJ. Corolla ontogeny and aestivation in the Acanthaceae. Bot J Linn Soc. 1994;114:49–65. https://doi.org/10.1111/j.1095–8339.1994.tb01923.x.
16. Scotland RW. Pollen morphology of Contortae (Acanthaceae). Bot JLinn Soc. 1993;111:471–504 http://ora.ox.ac.uk/objects/uuid:cd07aee0-b24c-4ef3-a5b8-816c0ae9e06c.
17. Scotland RW, Vollesen K. Classification of Acanthaceae. Kew Bull. 2000;55:513–89.
18. McDade LA, Thomas FD, Carrie AK, Vollesen K. Phylogenetic relationships among Acantheae (Acanthaceae): major lineages present contrasting patterns of molecular evolution and morphological differentiation. Syst Bot. 2005;30:834–62.
19. McDade LA, Thomas FD, Carrie AK. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae s.l. (Lamiales). Am. J. Bot. 2008;95:1136–52.
20. McDade LA, Thomas FD, Carrie AK, Agneta JB. Phylogenetic placement, delimitation, and relationships among genera of the enigmatic Nelsonioideae (Lamiales: Acanthaceae). Am J Bot. 2012;61:637–51.
21. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5:123–35 pmid:14735123.
22. Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber AP, et al. *Cyanophora paradoxa*genome elucidates origin of photosynthesis in algae and plants. Science. 2012;335:843–7 pmid:22344442.
23. Wei W, Youliang Z, Li C, Yuming W, Zehong Y, Ruiwu Y. PCR-RFLP analysis of cpDNA and mtDNA in the genus *Houttuynia* in some areas of China. Hereditas. 2005;142:24–32. https://doi.org/10.1111/j.1601-5223.2005.01,704.x.
24. Bendich AJ. Circular chloroplast chromosomes: the grand illusion. Plant Cell. 2004;16:1661–6. https://doi.org/10.1105/tpc.160771.
25. Asaf S, Khan AL, Khan MA, Imran QM, Kang SM. Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine otherGlycinespecies. PLOS ONE. 2017;12:e0182281. https://doi.org/10.1371/journal.pone.0182281.
26. Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, et al. Complete Chloroplast Genome Sequence of Tartary Buckwheat (*Fagopyrum tataricum*) and Comparative Analysis with Common Buckwheat (*F. esculentum*). PLOS ONE. 2015;10:e0125332. https://doi.org/10.1371/journal.pone.0125332.
27. Cronn R, Knaus BJ, Liston AP, Maughan J, Parks M, Syring J, et al. Targeted enrichment strategies for next-generation plant biology. Am J Bot. 2012;99:291–311. https://doi.org/10.3732/ajb.1100356.
28. Carstens BC, Pelletier TA, Reid NM, Satler JD. How to fail at species delimitation. Mol Ecol. 2013;22:4369–83. https://doi.org/10.1111/mec.12413.
29. Eaton D, Ree RH. Inferring phylogeny and introgression using RADseq data: an example from 347 flowering plants (*Pedicularis*: Orobanchaceae) Syst. Biol. 2013;62:689–706. https://doi.org/10.1093/sysbio/syt032.
30. Lemmon A, Emme S, Lemmon E. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst.Biol. 2012;61:727–44. https://doi.org/10.1093/sysbio/sys049.
31. Kamneva OK, Syring J, Liston A, Rosenberg NA. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. BMC Evol Biol. 2017;17. https://doi.org/10.1186/s12862-017-1019-7.
32. Sousa F, Bertrand YKJ, Nylinder S, Oxelman B, Eriksson JS, Pfeil BE, et al. Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation sequencing. PLOS ONE. 2014;9:e109704. https://doi.org/10.1371/journal.pone.0109704.
33. Stephens JD, Rogers WL, Heyduk K, Cruse-Sanders JM, Determann RO, Glenn TC, et al. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. Mol Phylogenet Evol. 2015;85:76–87.
34. Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, et al. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Appl Plant Sci. 2014;2:1400042. https://doi.org/10.3732/apps.1400042.
35. Clarke CB. Acanthaceea. In: Hooker JD, editor. The flora of British India, vol. 4. London: L. Reeve and Co.; 1884. p. 387–558.
36. Raman G, Park S. The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. Front Plant Sci. 2016;341.
37. Park I, Kim WJ, Yeo S-M, Choi G, Kang Y-M, Piao R, et al. The complete chloroplast genome sequences of *Fritillaria ussuriensis* maxim. In addition,

Alzahrani *et al. BMC Genomics*   (2020) 21:393

Page 18 of 19

Fritillaria cirrhosa D. Don, and comparative analysis with other *Fritillaria* species. Molecules. 2017;282..

38. Zhou J, Chen X, Cui Y, Sun W, Li Y, Wang Y. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. Int J Mol Sci. 2017;1839.

39. Jiang D, Zhao Z, Zhang T, Zhong W, Liu C, Yuan Q, et al. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. Genes. 2017;227.

40. Zhou J, Cui Y, Chen X, Li Y, Xu Z, Duan B, et al. Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: molecular structures, comparative analysis, and phylogenetic analysis. Molecules. 2018;437.

41. Wang W, Yu H, Wang J, Lei W, Gao J, Qiu X, et al. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (oleaceae) Int. J. Mol. Sci. 2017;2288.

42. Kumbhar F, Nie X, Xing G, Zhao X, Lin Y, Wang S, et al. Identification and characterisation of rna editing sites in chloroplast transcripts of einkorn wheat (*Triticum monococcum*). Ann Appl Biol. 2018;172:197–207.

43. Park M, Park H, Lee H, Lee B-H, Lee J. The complete plastome sequence of an antarctic bryophyte *Sanionia uncinata* (Hedw.) loeske. Int J Mol Sci. 2018;709.

44. Xumei W, Tao Z, Guoqing B, Yuemei Z. Complete chloroplast genome sequence of *Fagopyrum dibotrys*: genome features, comparative analysis and phylogenetic relationships. Sci Rep. 2018;8.

45. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol. 2001;16:142–7. https://doi.org/10.1016/S0169-5347(00)02097-8.

46. Philippe H, Delsuc F, Brinkmann H, Lartillot N. Phylogenomics. Ann Rev Ecol Evol Syst. 2005;36:541–62.

47. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boorem JL, et al. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics. 2007;8:174–201.

48. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM, et al. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. 2008;8:36–50.

49. Chunming G, Yunfei D, Jun W. The complete chloroplast genomes of *Echinacanthus* species (Acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. Front Plant Sci. 2019;1989.

50. Yongbin Z, Erin AT. The draft genome of *Ruellia speciosa* beautiful wild Petunia: Acanthaceae. DNA Res. 2017;24:179–92.

51. Ding P, Shao YH, Li Q, Gao JL, Zhang RJ, Lai XP, et al. The chloroplast genome sequence of the medicinal plant *Andrographis paniculata*. Mitochondr DNA. 2016;27:2347–8. https://doi.org/10.3109/19401736.2015.1025258.

52. Chen HM, Shao JJ, Zhang H, Jiang M, Huang LF, Zhang Z, et al. Sequencing and analysis of *Strobilanthes cusia* (Nees) Kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. Front Plant Sci. 2018;9. https://doi.org/10.3389/fpls.2018.00324.

53. Hildebrand M, Hallick RB, Passavant CW, Bourque DP. Trans-splicing in chloroplasts: the rps12 loci of *Nicotiana tabacum*. Proc Natl Acad Sci U S A. 1988;85:372–6. https://doi.org/10.1073/pnas.85.2.372.

54. Liu TJ, Zhang CY, Yan HF, Zhang L, Ge XJ, Hao G, et al. Complete plastid genome sequence of *Primula sinensis* (Primulaceae): struture comparison, sequence variation and evidence for *accD* tranfer to nucleus. PeerJ. 2016;4:e2101. https://doi.org/10.7717/peerj.2101.

55. Gichira AW, Li ZZ, Saina JK, Long ZC, Hu GW, Gituru RW, et al. The complete chloroplast genome sequence of an endemic monotypic genus *Hagenia* (Rosaceae): structural comparative analysis, gene content and microsatellite detection. PeerJ. 2017;5:e2846. https://doi.org/10.7717/peerj.2846.

56. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. Mol Biol Evol. 2013;31:645–59. https://doi.org/10.1093/molbev/mst257.

57. Lu L, Xiam L, Zhaodong H, Liming Y, Jingbo Z, Ye P, et al. Phylogenetic studies and comparative chloroplast genome analyses elucidate the basal position of halophyte *Nitraria sibirica* (Nitrariaceae) in the Sapindales. Mitochondr DNA Part A. 2017:1–11. https://doi.org/10.1080/24701394.2017.1350954.

58. Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G. Complete chloroplast genome of the multifunctional crop globe artichoke and comparisonwithotherAsteraceae. PLOS ONE. 2015;10:e0120589. https://doi.org/10.1371/journal.pone.0120589.

59. Bryan GJ, McNicol JW, Meyer RC, Ramsay G, De Jong WS. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor Appl Genet. 1999;99:859–67.

60. Provan J. Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. Mol Ecol. 2000;9:2183–5.

61. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour. 2009;9:673–90.

62. Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, et al. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. Front Plant Sci. 2017;8:1148. https://doi.org/10.3389/fpls.2017.01148.

63. Ye W-Q, Yap Z-Y, Li P, Comes HP, Qiu Y-X. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Mol Phylogenet Evol. 2018;127:978–87. https://doi.org/10.1016/j.ympev.2018.07.001.

64. 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;7:959. https://doi.org/10.3389/fpls.2016.00959.

65. Aldrich J, Cherney BW, Merlin E. The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. Curr Genet. 1988;14:137–46. https://doi.org/10.1007/bf00569337.

66. Zhou T, Chen C, Wei Y, Chang Y, Bai G, Li Z, et al. Comparative transcriptome and chloroplast genome analyses of two related Dipteronia species. Front Plant Sci. 2016;7. https://doi.org/10.3389/fpls.2016.01512.

67. Rousseau-Gueutin M, Bellot S, Martin GE, Boutte J, Chelaifa H, Lima O, et al. The chloroplast genome of the hexaploid *Spartina maritima* (Poaceae, Chloridoideae): comparative analyses and molecular dating. Mol Phylogenet Evol. 2015;93:516. https://doi.org/10.1016/j.ympev.2015.06.013.

68. Xu J-H, Liu Q, Hu W, Want T, Xue Q, Messing J. Dynamics of chloroplast genome in green plants. Genomics. 2015;106:221–31. https://doi.org/10.1016/j.ygeno.2015.07.004.

69. Borsch T, Quandt D. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Syst Evol. 2009;282:169–99.

70. Dong WP, Liu J, Yu J, Wang L, Zhou SL. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One. 2012;7:e35071.

71. Tong W, Kim TS, Park YJ. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. Rice. 2016;9:57.

72. Dong WP, Liu H, Xu C, Zuo YJ, Chen ZJ, Zhou SL, et al. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genet. 2014;15:138.

73. Du YP, Bi Y, Yang FP, Zhang MF, Chen XQ, Xue J, Zhang XH, et al. Complete chloroplast genome sequences of *Lilium*: insights into evolutionary dynamics and phylogenetic analyses. Sci Rep. 2017;7:5751.

74. Bentham G. Acanthaceae. In: Bentham G, Hooker JD, editors. Genera Plantarum, vol. 2:1060 – 1122. Reeeve, London Co., Lond; 1876.

75. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.

76. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2016;45.

77. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20:3252–5.

78. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 2005;33:686–W689.

79. 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;52:267–74.

80. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R, et al. Reputer: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.

81. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 2003;106:411–22.

82. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000;16:1046–7.

83. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:273–9.

84. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.

85. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–80.

86. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool. 1978;27:401–10.

87. Fredrik R, Maxim T, Paul VM, Daniel LA, Aaron D, Sebastian H, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic. 2012;61:539–42.

88. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008; 25:1253–9.

Publisher's Note

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