Physiol Mol Biol Plants (August 2021) 27(8):1867–1884
https://doi.org/10.1007/s12298-021-01051-w

RESEARCH ARTICLE

Complete chloroplast genome of the medicinal plant *Evolvulus alsinoides*: comparative analysis, identification of mutational hotspots and evolutionary dynamics with species of Solanales

P. R. Shidhi[1] · F. Nadiya[2] · V. C. Biju[1] · Sheethal Vijayan[1] · Anu Sasi[1] · C. L. Vipin[1] · Akhil Janardhanan[1] · S. Aswathy[1] · Veena S. Rajan[1] · Achuthsankar S. Nair[1]

Received: 19 January 2021 / Revised: 13 August 2021 / Accepted: 16 August 2021 / Published online: 25 August 2021

Abstract *Evolvulus alsinoides*, belonging to the family Convolvulaceae, is an important medicinal plant widely used as a nootropic in the Indian traditional medicine system. In the genus *Evolvulus*, no research on the chloroplast genome has been published. Hence, the present study focuses on annotation, characterization, identification of mutational hotspots, and phylogenetic analysis in the complete chloroplast genome (cp) of *E. alsinoides*. Genome comparison and evolutionary dynamics were performed with the species of Solanales. The cp genome has 114 genes (80 protein-coding genes, 30 transfer RNA, and 4 ribosomal RNA genes) that were unique with total genome size of 157,015 bp. The cp genome possesses 69 RNA editing sites and 44 simple sequence repeats (SSRs). Predicted SSRs were randomly selected and validated experimentally. Six divergent hotspots such as *trnQ-UUG*, *trnF-GAA*, *psaI*, *clpP*, *ndhF*, and *ycf1* were discovered from the cp genome. These microsatellites and divergent hot spot sequences of the Taxa '*Evolvulus*' could be employed as molecular markers for species identification and genetic divergence investigations. The LSC area was found to be more conserved than the SSC and IR region in genome comparison. The IR contraction and expansion studies show that nine genes *rpl2*, *rpl23*, *ycf1*, *ycf2*, *ycf1*, *ndhF*, *ndhA*, *matK*, and *psbK* were present in the IR-LSC and IR-SSC boundaries of the cp genome. Fifty-four protein-coding genes in the cp genome were under negative selection pressure, indicating that they were well conserved and were undergoing purifying selection. The phylogenetic analysis reveals that *E. alsinoides* is closely related to the genus *Cressa* with some divergence from the genus *Ipomoea*. This is the first time the chloroplast genome of the genus *Evolvulus* has been published. The findings of the present study and chloroplast genome data could be a valuable resource for future studies in population genetics, genetic diversity, and evolutionary relationship of the family Convolvulaceae.

Keywords *Evolvulus alsinoides* · Chloroplast genome · Genes · Simple sequence repeats · Purifying selection

Abbreviations

AA	Amino Acid
Cp	Chloroplast
IR	Inverted repeat
LSC	Large Single Copy
PCGs	Protein Coding Genes
RSCU	Relative Synonymous Codon Usage
SMRT	Single Molecule Real Time
SSC	Small Single Copy
SSR	Simple Sequence Repeat

Introduction

E. alsinoides, commonly known as shankhpushpi belonging to the Convolvulaceae family, is an essential plant in the Indian traditional medicine system for treating various ailments (Sethiya et al. 2010). It can be seen in regions with tropical and subtropical climates. The plant is extensively found in Kerala's western Ghats region, where it is one of

✉ P. R. Shidhi
shidhibio@gmail.com

[1] Department of Computational Biology and Bioinformatics, University of Kerala, Thiruvananthapuram, Kerala, India

[2] Department of Biotechnology, Inter University Centre for Genomics and Gene Technology, University of Kerala, Thiruvananthapuram, Kerala, India

the ten sacred plants known as 'Dashapushpam' (Austin 2008; Priya 2017; Uthaman and Nair 2017). Jeevanyadi ghrita, Brahma rasayana, Vachadi ghrita, Brahmi ghrita, Naladi ghrita and Agastya rasayana are renowned preparations in Ayurveda that have Dashapushpam as the vital ingredient (Madhavan et al. 2008). The various phytochemicals present in the plant possess antibacterial, antifungal, antiviral, anticancer, neurodegenerative, antidiabetic, antioxidant, and immunomodulatory activities (Priya 2017; Ambika and Nair 2019; Biémont and Vieira 2006; Sethiya et al. 2019; Singh 2008; Siraj et al. 2019). The green plants have chloroplast, which is a vital organelle responsible for photosynthesis, carbon fixation, biosynthesis of amino acids, fatty acids, starch, and pigments (Jansen et al. 2007; Kim et al. 2015; Sugiura 1992). Chloroplast DNAs (cpDNAs) are double-stranded molecules, existing in a circular model, ranging it's size from 35 to 217 Kbp, but the size varies from 115 to 165 Kbp in terrestrial plants, with a structure that is quadripartite comprising small single copy (SSC) and large single copy (LSC) region that is parted by two inverted repeats (Sugiura 1992, 2005; Ravi et al. 2008; Wicke et al. 2011). The chloroplast genome (cp genome) has been categorized into introns, protein-coding genes, and intergenic spacers based on functions. The introns and intergenic spacers that do not encode proteins are referred to as non-coding regions. It could be used for phylogenetic, molecular evolution, and population genetic studies (Shaw et al. 2007). The first complete cpDNA was sequenced in 1986 for *Marchantia polymorpha*; after that, several other cpDNA have been sequenced and characterized from several plants to understand the evolutionary and functional information (Supriya and Priyadarshan 2019). Next-generation sequencing techniques allow the generation of massive sequence data at a low cost. It has paved the way for molecular ecology research which has made cp genome sequencing more feasible (Henry et al. 2014). The cp genome sequencing has developed as an alternative solution for marker identification, DNA barcoding, species identification, and phylogenetic studies. In general, the nuclear genome is less conserved than cp genome (Kim et al. 2015; Park et al. 2018; Erixon and Oxelman 2008; Erixon and Oxelman 2008; Sloan et al. 2014).

Although cp genome of other genera belonging to the Convolvulaceae family has been sequenced and used in phylogenetic studies, no cp genome of the genus *Evolvulus* has been sequenced and published so far. Hence, in the current study, the complete cp genome of *E. alsinoides* was sequenced and characterized using Illumina and PacBio sequencing technologies. The denovo assembled sequence was annotated and compared with the previously reported cp genome of Solanales to understand the gene content, genome organization, and its evolutionary history.

Materials and methods

Plant material and DNA extraction

The plant *E. alsinoides* used for the study was gathered from Kariavattom North campus, 8.5674° N; 76.8879° E, elev. 60 m. University of Kerala, Trivandrum, Kerala, India. Under the voucher number KUBH9910, the voucher specimen was deposited at the Herbarium of the Department of Botany, University of Kerala, India. CTAB (cetyltrimethylammonium bromide) method (Healey et al. 2014) was used to isolate whole genomic DNA from leaf tissues. Nanodrop (Nanodrop Technologies, Wilmington, DE, US) and 1% (w/v) agarose gel electrophoresis were used to estimate the purity of the sample, and the quantification was done using the Qubit system (Thermo Fisher Scientific, Waltham, MA, USA).

Library preparation and genome sequencing

A total amount of 1 μg of high-quality genomic DNA was used for library preparation. The genomic DNA was randomly fragmented to 300 bp using the S220 Focused-ultrasonicator system (Covaris, Woburn, MA, USA). The NEBnext Ultra DNA library prep kit for Illumina (E7370L; New England Biolabs, Ipswich, MA, USA) was used to prepare sequencing libraries as per manufacturer's instructions. Paired-end sequencing was done in Illumina HiSeqX platform (2 × 150 bp; Illumina, San Diego, CA, USA). For PaCBio sequencing, 5 μg of mechanically sheared genomic DNA [using Covaris g-TUBE (Covaris, Woburn, MA, USA)] of size range 10–20 Kb were enzymatically repaired and converted into Single-Molecule Real-Time bell templates (SMRTbell) according to the manufacturer's protocol. The SMRTbell template libraries were prepared using SMRTbell Template Prep Kit 1.0 (100-259-100; Pacific Biosciences, Menlo Park, CA, USA). The resulting SMRTbell templates were then size-selected by Blue Pippin electrophoresis (Sage Sciences). The SMRTbell templates ranging from 15 to 20 Kb were sequenced on a PacBio Sequel Sequencing chemistry 2.1 (Pacific Biosciences, Menlo Park, CA, USA) instrument with 3 SMRT cells.

Assembly and annotation

The FastQC was used to perform the quality of the raw reads (v0.10.0) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The preprocessing step was performed using high quality Illumina raw reads. Cutadapt v1.8 (Martin 2011) were used to remove the adaptor sequences, and Sickle v 1.33 (https://github.com/najoshi/sickle) were

used to perform quality trimming of the reads. To extract the chloroplast reads, preprocessed Illumina and PacBio reads were compared to a reference set of previously published cp genomes from 5 Convolvulaceae species obtained from the National Center for Biotechnology Information (NCBI) database (Supplementary Table S1). The reference genome set was mapped to Illumina reads using Bowtie2 v2.2.6 (Langmead and Salzberg 2012), with default parameters, and PacBio reads were aligned to the reference genome set using minimap2-2.17 (Li 2018). All mapped PacBio reads were extracted, filtered, and the error rectified with aligned Illumina reads using Proovread v2.12 (Hackl et al. 2014) and Canu v1.8 was used to assemble the resulting data (Koren et al. 2017). The raw reads were remapped to the final cp genome assembly using Bowtie2 v2.2.6. to verify the coverage and quality of the assembled genome. Genome annotation was carried out using GeSeq (Tillich et al. 2017) with reference species of the family Convolvulaceae (Supplementary Table S1). The protein-coding regions were confirmed using BLAST (Altschul et al. 1990) analysis. The codon positions (start and stop) and the borders of introns and exons, were manually adjusted using the cp genome reference sets. Predicted tRNAs were confirmed by tRNAscan-SE 1.21 (http://lowelab.ucsc.edu/tRNAscan-SE/) (Chan and Lowe 2019; Lowe and Chan 2016). Organellar Genome DRAW (OGDRAW) (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Lohse et al. 2013) was used to generate a map of the genome with a circular shape.

Codon usage analysis and RNA editing site prediction

CodonW v1.4.4 (http://codonw.sourceforge.net/) was used to carry out synonymous codon usage analysis of the protein-coding genes. PREPsuite (http://prep.unl.edu/) (Du et al. 2009) was used to identify the RNA editing sites in the protein-coding region of *E. alsinoides* with default settings.

Analysis of microsatellites and oligonucleotide repeats

Simple sequence repeats (SSRs) referred as microsatellites were identified using the MISA-Web (https://webblast.ipk-gatersleben.de/misa/) (Beier et al. 2017) with the default parameters. A comparison of SSRs was also performed among the cp genome of 5 species belongs to the family Convolvulaceae (Supplementary Table S1). The minimum number of repeats used for the analysis was set as default. REPuter v2.74 (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) (Kurtz et al. 2001) was used to identify four different kinds of repeats, such as forward repeat, reverse repeat, complement repeat, and palindromic repeat.

Experimental validation of SSR markers

DNA isolation was performed using above mentioned method. SSR primers flanking the microsatellites were designed using Primer3 (https://primer3.ut.ee/) (Untergasser et al. 2012) and Perl script provided by the MISA (https://webblast.ipk-gatersleben.de/misa/) (Beier et al. 2017). Fifteen SSR primer pairs were synthesized, and gradient PCR was performed to determine the annealing temperature. Different temperatures variations were checked for each primer. DNA amplification was performed in a 25 µL reaction mixture containing 1 × Buffer (NEB, USA), 200 µM dNTPs (NEB, USA), 1 µL of two primers (forward and reverse primers) (10 Pico Moles, Europhins, Luxembourg), 1U Taq DNA polymerase (NEB, USA) and 50 ng DNA template. SSR-PCR amplification was performed in a C1000 Touch Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA) with the following program: initial denaturation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 40 s, primer annealing with specific temperature for 40 s, 72 °C for 1 min 30 s, final extension for 5 min at 72 °C. On a 3% agarose gel the amplified products were separated, stained with ethidium bromide, visualized and photographed using Chemi Doc image documentation analysis system (BioRad, USA). Fragment sizes were estimated with 1 kb DNA Ladder sizing markers (NEB, USA). A similarity search was executed with amplified SSRs using BLAST with an e-value of 1E−5 to identify the functional homologs.

Genome comparison, sequence variation, IR contraction, and expansion

Genome comparison was performed using cp genome of *E. alsinoides* and 5 species, which are closely related to the family Convolvulaceae representing three genera (*Cressa, Ipomoea*, and *Cuscuta*) (Supplementary Table S1). Multiple sequence alignment was performed using mVISTA program (available online: http://genome.lbl.gov/vista/index.shtml) which compares the cp genome of *E. alsinoides* with other selected species using Shuffle-LAGAN mode to estimate the divergence of sequence. Using the IRscope tool (https://irscope.shinyapps.io/irapp/), a comparison of junction sites among small single copy (SSC), inverted repeat (IR), and large single copy (LSC) were done.

Synonymous and non-synonymous substitutions rates and divergent hotspot identification

The individual protein-coding genes (Katoh and Standley 2013) MAFFT v7 (https://mafft.cbrc.jp/alignment/server/) were used to map the individual protein-coding genes of *E. alsinoides* and 18 species of the order Solanales, and translated to protein sequence. DNAsp v5.10.01(Rozas et al. 2003) was used to identify nonsynonymous (Ka) and synonymous (Ks) substitution rate and Ka/Ks ratio to identify the genes under selection pressure. The above sequence alignment was used to calculate the nuclear diversity (Pi) using DnaSP v5.10.01 software (Rozas et al. 2003).

Phylogenetic analysis and divergence time estimation

For phylogenetic analysis, Convolvulaceae and Solanaceae families were included within the order Solanales as among the families, Solanaceae is closely related to Convolvulaceae. The cp genomes of 33 species were selected (Supplemental Table S2). The in-group contains the genomes of 32 species from the order Solanales, including fourteen Convolvulaceae (five *Cuscuta* species, seven *Ipomoea* species, *C. cretica,* and *E. alsinoides*) and eighteen Solanaceae (five *Capsicum* species, six *Solanum* species, and six *Nicotiana* species and *Petunia exserta*). *Arabidopsis thaliana* was selected as the outgroup since it is a well-studied model organism and it has been used as an outgroup in several studies involving identification, characterization, and annotation of organelle genome (Zhao et al. 2019; De-la-Cruz and Núñez-Farfán 2020). Sequences of all the cp genomes were downloaded from the NCBI Organelle Genome Resources database. Conserved protein-coding genes were extracted using local Perl script and aligned using MAFFT 7 (Katoh and Standley 2013). The alignment was adjusted manually and concatenated to get a super alignment for constructing the phylogenetic tree. The best substitution model for constructing phylogenetic trees was predicted using jModelTest version 3.7. Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) were employed for creating the phylogenetic tree. The ML analysis was performed by MEGA v7.0.26 (Kumar et al. 2016) with 1000 bootstrap replicates. The MP analysis were done using PAUP* v4.0a165 (Wilgenbusch and Swofford 2003). 1000 replicates of random taxon addition, tree bisection-reconnection branches wrapping, multitree on, the collapse of zero-length branches, and numerous tree options were used in a heuristic search, with each replicate saving a maximum of 100 trees. BI analysis was performed using the MrBayes program v3.2.7a (Huelsenbeck and Ronquist 2005).

BEAST v1.10.1 (Yoder and Yang 2000; Li and Drummond 2012) was used to estimate the divergence times based on the Bayesian method with the GTR + GAMMA substitution model, with a strict molecular clock and Yule model. The tree model used was a random starting tree. The super alignment was used to analyze the divergence time. The calibration points for the species (divergence between *Ipomoea nil* and *Arabidopsis thaliana* was 121 million years ago) were obtained from Tree time webserver (Kumar et al. 2017), based on previously published calibration time. BEAST Monte Carlo Markov Chain (MCMC) simulations were run for 10,000,000 generations to generate the tree file. The tree file was annotated using Tree Annotator v1.6.1 software. FigTree v1.4.3 was used to create the final tree with node ages.

Results and discussion

Assembly and annotation

A total of 496,329,876 Illumina (74 Gb) and 3,576,937 PacBio (20.99 Gb) raw reads were generated after sequencing of leaf tissues (SRA ID: PRJNA551197). The pre-processing step resulted in 450,360,918 cleaned high quality Illumina reads. Out of these, 38,132,788 (7.68% of total reads) Illumina reads and 203,338 (5.68% of total reads) PacBio reads were mapped to the whole cp genome of closely related species. The de novo assembly using error corrected high-quality Illumina reads and PacBio reads created a circular cp genome of *E. alsinoides* with a size of 157,015 bp. With the entry number MN548282, the whole cp genome sequence of E. alsinoides has been deposited in the NCBI GenBank database. The cp genomes of all angiosperm possess a typical quadripartite structure with an LSC region, SSC region, and repeat regions (IRa/IRb) (Wicke et al. 2011; Jones and Kang 2015). The LSC region (85,820 bp) and SSC region (13,691 bp) of *E. alsinoides*' cp genomes are separated by a couple of inverted repeats, IRa and IRb (28,752 bp) (Fig. 1). The junction regions between IR, LSC, and SSC of the cp genome may vary among all angiosperms (Wang et al. 2013; Xu et al. 2012; Zhang et al. 2011). The cp genome of *E. alsinoides* consists of 114 unique genes, including 80 PCGs, 30 tRNAs, and 4 rRNAs (Table 1). The SSC region has 11 PCGs and 1tRNA, while LSC region has 63 PCGs and 23tRNAs (Supplementary Table S3). Of 114 genes, 6 PCGs (*ycf1*, *ycf2*, *ycf15*, *rps7*, *rps12*, *ndhB*), 6 tRNA genes (*trnL-CAA*, *trnA-UGC*, *trnI-GAU*, *trnN-GUU*, *trnV-GAC*, *trnR-ACG*) and 4 rRNA genes (rrn16, rrn23, rrn4.5, rrn5) in the IR regions were duplicated. The cp genome of *E.*

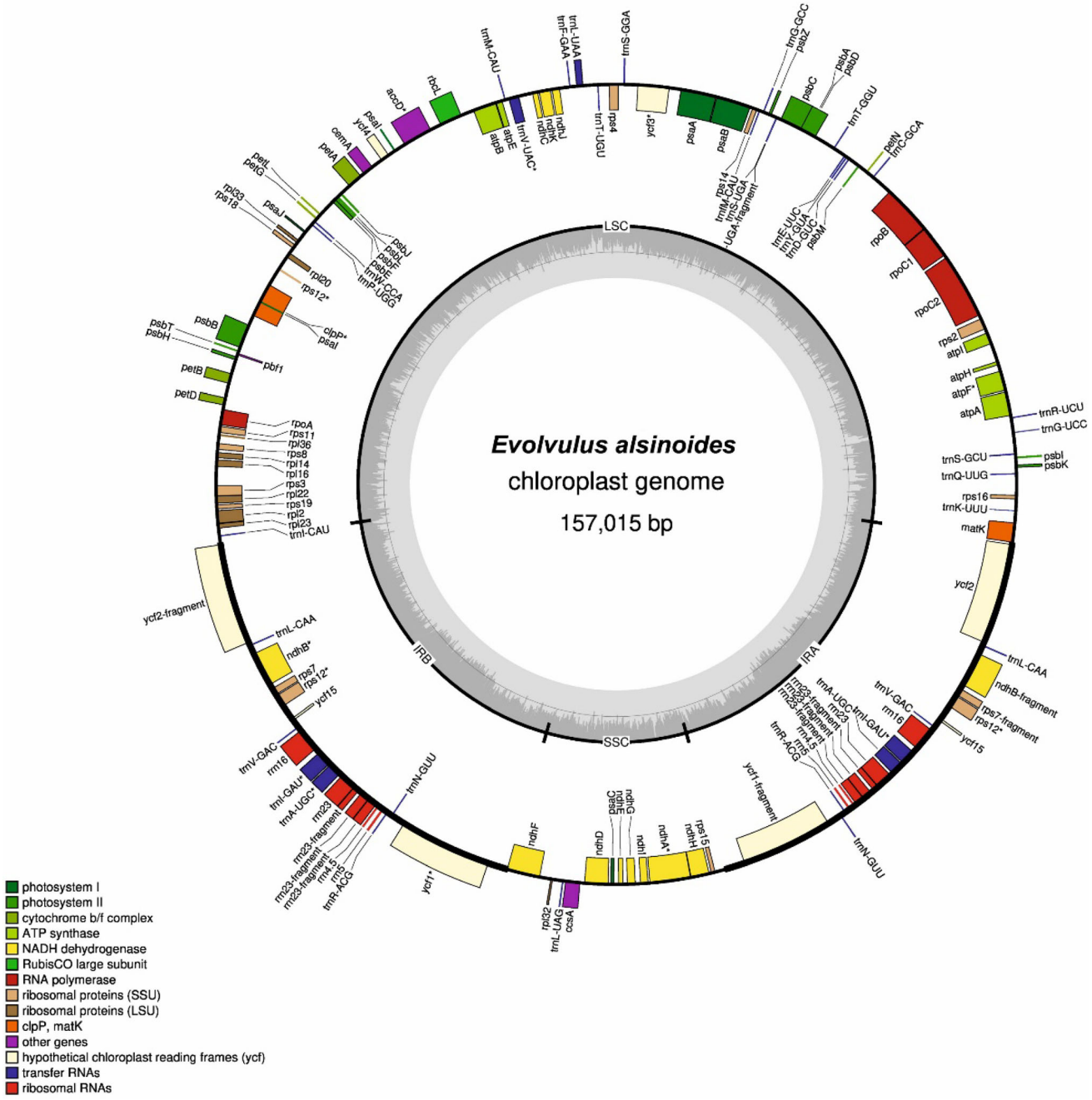

Fig. 1 cp genome map of *E. alsinoides*. Genes in the inner circle are transcribed clockwise, and genes in the outer circles are transcribed in counter-clockwise direction. The IR/LSC/SSC boundaries are represented in the inner circle. The inner grey circle represent the GC and AT content of the cp genome

alsinoides contains 13 intron-containing genes [7 PCGs and 6 tRNA (Table 1)]. The trans-spliced gene 'rps12' is seen in the LSC region's 5′ end and repeated in the IR region's 3′ end. The *matK* gene was positioned inside the intron of *trnK-UUU*. In the border zones between IR and SSC, one pseudogene (*ycf1*) was found. The GC content of the cp genome was 37.3%, which was similar to the published cp genome of other species belonging to the family Convolvulaceae (Funk et al. 2007; Park et al. 2018, 2019). IR regions had a GC content of 41.2%, which was higher than the GC content of LSC (39.5%) and SSC (32.2%) regions.

Codon usage analysis

In plants, codon usage bias refers to synonymous codons having different usage frequencies. Several evolutionary mechanisms that may lead to gene mutation and selection

Table 1 Genes present in *E. alsinoides* cp genomes

Gene category	Gene group	Gene name
Photosynthesis related genes	Photosystem I	*psaA*, *ycf4*, *psaI*, *psaB*, *psaC*, *ycf3*[c], *psaJ*
	Photosystem II	*psbH*, *psbI*, *psbA*, *psbD*, *psbB*, *psbC*, *psbF*, *psbE*, *psbJ*, *psbT*, *psbZ psbK*, *psb*, *psbN*, *pbf1*, *psbM*
	Cytochrome b/f complex	*petB*[b], *petD*[b], *petA*, *petG*, *petL*, *petN*
	ATP synthase	*atpH*, *atpA atpE*, *atpB*, *atpF*[b], *atpI*
	NADH dehydrogenase	*ndhG*, *ndhA*[b], *ndhE*, *ndhD*, *ndhC*, *ndhB*[ab], *ndhH*, *ndhF*, *ndhI*, *ndhK*, *ndhJ*
Transcription and translation-related genes	RNA polymerase	*rpoC2*, *rpoA*, *rpoC1*[b], *rpoB*
	Large subunit ribosomal proteins	*rpl14*, *rpl2*, *rpl16*, *rpl20*, *rpl23*, *rpl32*, *rpl22*, *rpl36*, *rpl33*
	Small subunit ribosomal proteins	*rps2*, *rps8*, *rps4*, *rps7*[a], *rp3*, *rps14*, *rps12*[abc], *rps15*, *rps11*, *rps16*[b], *rps19*, *rps18*
RNA genes	Ribosomal RNA genes	*rn4.5*[a], *rrn23*[a], *rrn5*[a], *rrn16*[a]
	Transfer RNA genes	*trnA-UGC*[ac], *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG- GCC*, *trnG-UCC*, *trnI-CAU*, *trnfM-CAU*, *trnK-UUU*, *trnI-GAU*[ac], *trnL-CAA*[a], *trnL-UAA*[b], *trnL-UAG*, *trnM-CAU*, *trnN-GUU*[a], *trnP-UGG*, *trnR- ACG*[a], *trnQ-UUG*, *trnR-UCU*, *trnS-CGA*[b], *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*[a], *trnV-UAC*[b], *trnW- CCA*, *trnY-GUA*
Other genes	Fatty acid synthesis	*accD*[b]
	Carbon metabolism	*cemA*
	Cytochrome c synthesis	*ccsA*
	Proteolysis	*clpP*[c]
	RNA processing	*matK*
	Rubisco	*rbcL*
Genes of unknown function	Conserved reading frames	*ycf1*[ab], *ycf2*[a], *ycf15*[a]

[a]Similar gene in IRs

[b]Single intron gene

[c]Two intron gene

influenced the occurrence of codon bias (Ermolaeva 2001; Wong et al. 2002). Usage patterns of Codon may differ among different genes and species. An adequate number of codons (Nc) of 80 PCGs of *E. alsinoides* were calculated to study codon usage analysis (Supplementary Table S4). Results show that Nc values varied between 30.55 (*rpl36*) and 61 (*psbM*) in all the identified PCGs. The majority of Nc values in the *E. alsinoides* cp genome were greater than 44, indicating a fragile gene codon bias. With a mean Nc value of 30.55, the *rpl36* gene was found to have the most biassed codon use. (Supplementary Table S4). The 79 unique PCGs consisted of 66,448 bp that encoded 22,026 codons. The most frequently used AA in the *E. alsinoides* cp genome was leucine, which was encoded by 2366 (10.74%) of these codons (Table 2). Only 312 (1.42%) codons encode cysteine, the least frequent amino acid. It was reported that in angiosperms leucine is the frequently used AA while cysteine is the rarely used AA. (Chen et al. 2015; Dong et al. 2012; Kaila et al. 2016). The Relative Synonymous Codon Usage was used to calculate the codon usage bias (Sharp and Cowe 1991). The frequency of a codon divided by the expected frequency gives Relative Synonymous Codon Usage. In the *E. alsinoides* cp genes, thirty codons showed codon use bias due to RSCU values of > 1. Twenty-nine of the thirty codons had A (adenine) or U (uracil) endings. The G (guanine) or C (cytosine) ending codons (RSCU values 1) showed the contrary pattern, indicating that they are less prevalent in *E. alsinoides* cp genes. The use of stop codons was discovered to be skewed toward UAA. Poplar, rice, and other plants were found to have similar codon usage rules of bias for A (adenine) or T

Table 2 The Relative synonymous codon usage of the *E. alsinoides* cp genome

Amino acid	Codon	Number	RSCU	AA frequency	Amino acid	Codon	Number	RSCU	AA frequency
Phe	UUU	857	**1.33**	5.86	Tyr	UAU	630	**1.61**	3.56
	UUC	434	0.67			UAC	155	0.39	
Leu	UUA	721	**1.83**	10.74	TER	UAA	106	**1.39**	1.04
	UUG	494	**1.25**			UAG	56	0.73	
	CUU	480	**1.22**			UGA	67	0.88	
	CUC	142	0.36						
	CUA	365	0.93		His	CAU	383	**1.46**	2.37
	CUG	164	0.42			CAC	140	0.54	
Ile	AUU	923	**1.48**	8.49	Gln	CAA	595	**1.43**	3.79
	AUC	382	0.61			CAG	239	0.57	
	AUA	565	0.91						
					Asn	AAU	695	**1.47**	4.28
Met	AUG	517	1.00	2.35		AAC	249	0.53	
Val	GUU	427	**1.43**	5.42	Pro	CCU	324	**1.44**	4.09
	GUC	180	0.6			CCC	205	0.91	
	GUA	420	**1.41**			CCA	244	**1.08**	
	GUG	167	0.56			CCG	129	0.57	
Lys	AAA	886	**1.48**	5.45	Trp	UGG	368	1	1.67
	AAG	315	0.52						
					Ala	GCU	525	**1.74**	5.47
Cys	UGU	220	**1.41**	1.42		GCC	213	0.71	
	UGC	92	0.59			GCA	329	**1.09**	
						GCG	139	0.46	
Ser	UCU	418	**1.54**	5.31					
	UCC	269	0.99		Glu	GAA	791	**1.49**	4.83
	UCA	317	**1.17**			GAG	273	0.51	
	UCG	165	0.61						
					Thr	ACU	425	**1.54**	5
Ser	AGU	345	**1.27**	2.07		ACC	208	0.75	
	AGC	111	0.41			ACA	342	**1.24**	
						ACG	128	0.46	
Arg	AGA	400	**1.76**	2.49					
	AGG	149	0.66		Asp	GAU	660	**1.58**	3.8
						GAC	176	0.42	
Arg	CGU	252	**1.11**	3.7					
	CGC	105	0.46		Gly	GGU	472	**1.26**	6.78
	CGA	327	**1.44**			GGC	161	0.43	
	CGG	129	0.57			GGA	556	**1.49**	
						GGG	305	0.82	

The significance of relative synonymous codon usage (RSCU) > 1 is denoted in bold

(thymine) (Shaw et al. 2007; Wang et al. 2013; Qingpo and Qingzhong 2004).

RNA editing sites prediction

E. alsinoides cp genome contains 69 RNA editing sites, which was distributed among 20 PCGs (Table 3). The organelle genome of most plants has RNA editing phenomenon (Bock 2000). Out of 69 RNA editing sites, 33 are

Table 3 The predicted RNA editing site in the *E. alsinoides* cp genes

Gene	Codon conversion	Amino acid conversion	Gene	Codon conversion	Amino acid conversion
rps16	CTT = > TTT	L = > F	*ycf3*	ACT = > ATT	T = > I
atpI	TCA = > TTA	S = > L			
rps2	TCA = > TTA	S = > L	*accD*	GCA = > GTA	A = > V
rpoC2	CAT = > TAT	H = > Y		GCT = > GTT	A = > V
	CAT = > TAT	H = > Y		CTT = > TTT	L = > F
	CAC = > TAC	H = > Y			
	CGG = > TGG	R = > W	*psaI*	TCA = > TTA	S = > L
	CGC = > TGC	R = > C			
	CCG = > TCG	P = > S	*psbL*	CTT = > TTT	L = > F
rpoC1	ACA = > ATA	T = > I			
	CCC = > TCC	P = > S	*psbF*	CCT = > CTT	P = > L
	ACT = > ATT	T = > I		CTT = > TTT	L = > F
	TCG = > TTG	S = > L			
	CCT = > CTT	P = > L	*psbE*	CCT = > TCT	P = > S
	CAT = > TAT	H = > Y			
	ACA = > ATA	T = > I	*petG*	TCT = > TTT	S = > F
rpoB	TCA = > TTA	S = > L	*psaI*	TCA = > TTA	S = > L
	TCA = > TTA	S = > L			
	TCG = > TTG	S = > L	*psbB*	CTT = > TTT	L = > F
	TCT = > TTT	S = > F		CTT = > TTT	L = > F
	GCA = > GTA	A = > V		CTT = > TTT	L = > F
				CTT = > TTT	L = > F
rps14	TCA = > TTA	S = > L		CCT = > CTT	P = > L
	CCA = > CTA	P = > L		CTT = > TTT	L = > F
				CTT = > TTT	L = > F
psaB	CTT = > TTT	L = > F		CTT = > TTT	L = > F
	CTT = > TTT	L = > F		CTT = > TTT	L = > F
	CCT = > CTT	P = > L		CTT = > TTT	L = > F
	CTT = > TTT	L = > F		CTT = > TTT	L = > F
	CCG = > TCG	P = > S		CCT = > CTT	P = > L
	CCT = > CTT	P = > L			
	CCT = > CTT	P = > L	*petB*	CCA = > CTA	P = > L
	CTT = > TTT	L = > F		ACC = > ATC	T = > I
	CTT = > TTT	L = > F	*petD*	CTT = > TTT	L = > F
	CTT = > TTT	L = > F		CTT = > TTT	L = > F
	CTT = > TTT	L = > F		CCT = > CTT	P = > L
	ACA = > ATA	T = > I		CCT = > CTT	P = > L
	CAC = > TAC	H = > Y			
	TCT = > TTT	S = > F	*rpoA*	CTT = > TTT	L = > F
	CCT = > CTT	P = > L			
	CTT = > TTT	L = > F			

C- to -T conversion which is the most common type in Organellar transcripts of plants (Jansen et al. 2007; Duruvasula et al. 2019; Maier et al. 1995). The *psaB* (16) gene has maximum RNA editing sites trailed by *psbB* (12), *rpoC1* (7), *rpoC2* (6), *rpoB* (5), *petD* (4), *accD* (3), *psaI, psbF, petB, rps14* (2 each), *rps16*, *atp1, rps2*, *ycf3*, *psbL*, *psbE*, *petG*, *rpoA* Meanwhile, the genes *ndhD* and *rpoB* (1 each). The highest transition in cp genome is the Leucine to Phenylalanine, trailed by Proline to Leucine and Serine to Leucine. The location of RNA editing sites present in the

cp genome was 1st codon (36), 2nd codon (33), and 3rd codon (0), respectively (Supplementary Table S5). The RNA prediction site in the first codon of the first nucleotide was absent in the genes such as *atp1*, *rps2*, *rpoB*, *rps14*, *ycf3*, *accD*, *PsaI*, and *PetG*. The following genes *atp1*, *rps2*, *rpoB*, *rps14*, *ycf3*, *accD*, *PsaI*, and *PetG* do not have an RNA predicting site in their first nucleotide codon.

Analysis of microsatellites and oligonucleotide repeats

Microsatellites also referred as SSRs are short nucleotides stretches that are widely present in the genomes. 49 SSRs were identified in the cp genome of *E. alsinoides*. Out of it, the majority were mononucleotide repeats (44) trailed by dinucleotide (1), trinucleotide (2) tetranucleotide (1) and hexanucleotide (1) repeats (Fig. 2A). Of the 49 SSRs, 88% of the mononucleotide SSRs were (A/T) Poly T (poly-thymine) or poly-A (polyadenine) repeats in the cp genome of *E. alsinoides* (Fig. 2B) similar to preceding studies of other species (Saina et al. 2018; Zhou et al. 2016; Qian et al. 2013) (Supplementary Table S6). About 42 SSRs were situated in the non-coding regions (IGS intergenic spacers) and 7 SSRs, including *ycf3*, *accD*, *clpP* and *ndhA* were located in the genic region. The majority of SSRs were found in the LSC region with 73%, and the SSC and IR regions contain 22% and 5% respectively. Two mononucleotides are present in the IR region, which was the duplicated SSR. In total, 44 non-redundant SSRs were present in the cp genome of *E. alsinoides.* SSRs are used as molecular indicators in genetic heterogeneity studies, plant breeding programs, evolutionary studies, and species identification (Chan and Lowe 2019; Bryan et al. 1999; Ebert and Peakall 2009; Ebert and Peakall 2009; Yang et al. 2016; Powell et al. 1995; Dong et al. 2013). A comparison of SSRs was performed among the cp genome of 5 species belongs to the family Convolvulaceae (Supplementary Table S6). Among all the species, *E. alsinoides* and *Ipomoea purpurea* have the maximum number of SSRs, and *Cuscuta exaltata* has the least number of SSRs (Fig. 3). The length of the repeats in all the species ranges from 10 to 115 bp. *I. nil* and *I. purpurea* have only mononucleotide repeats. *Cuscuta reflexa* and *C. exaltata* have mono, and di nucleotide repeats, whereas *Cressa*

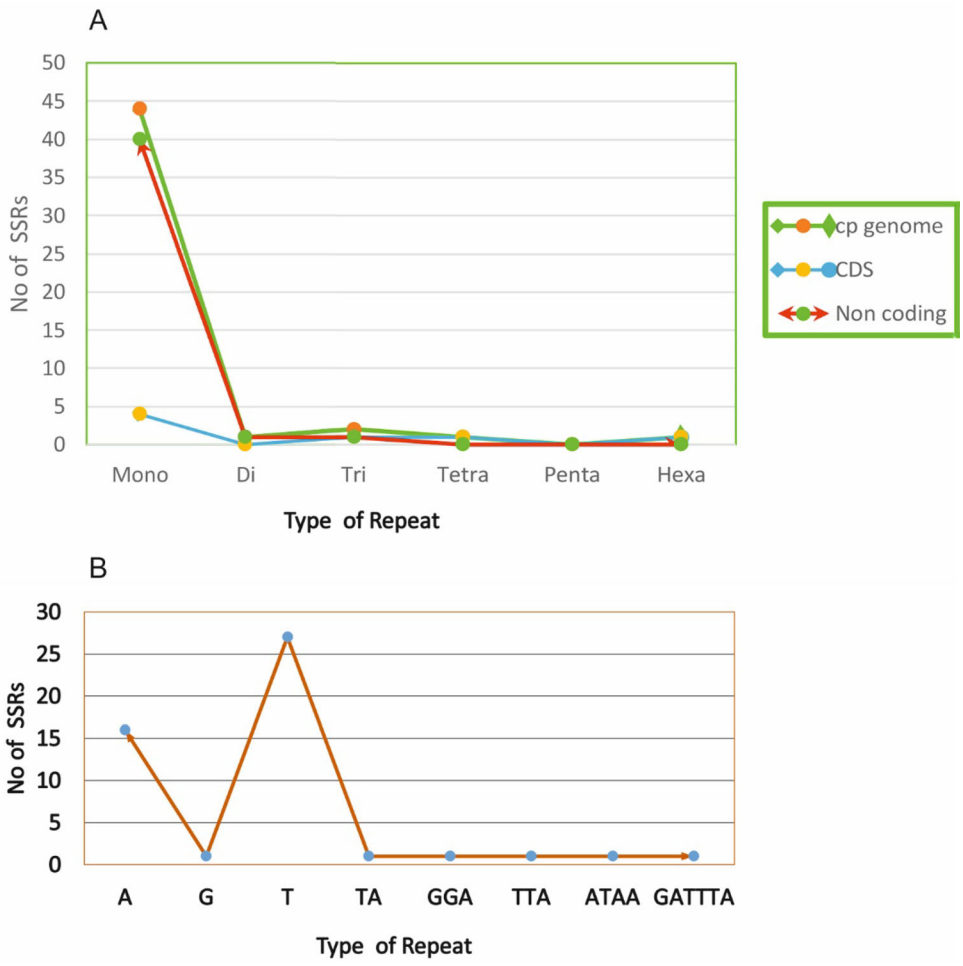

Fig. 2 **A** Number of SSRs in cp genome, coding region and non-coding region. **B** Frequency of SSR motifs in the cp genome. The X-axis represent the type of repeats and the Y-axis represents the No of SSRs

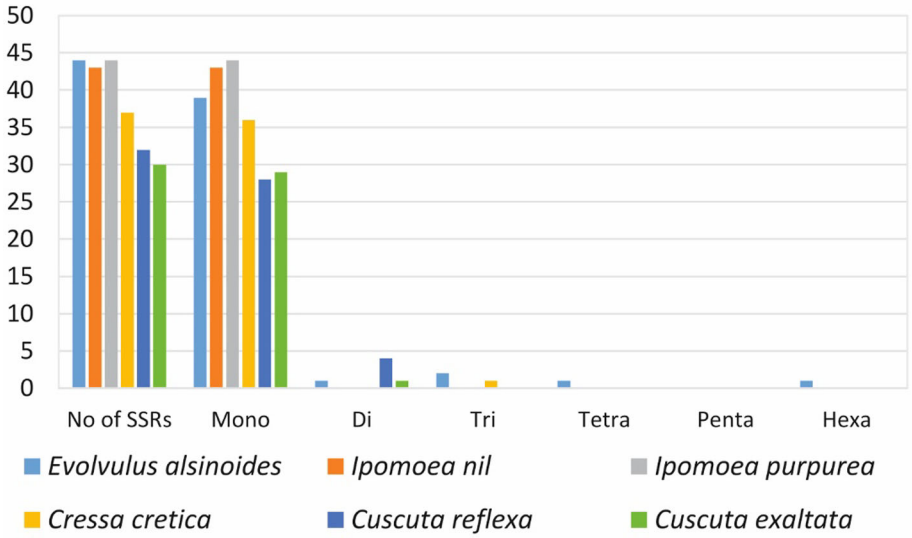

Fig. 3 Comparison of SSRs with the closely related species of *E. alsinoides*

cretica has mono and tri nucleotide repeats. *C. cretica* and *I. purpurea* have the shortest length of repeats. The A/T mono- repeats were profusely existing in all the species chloroplast genome used for the study.

In addition to the SSRs, repeat analysis identified 86 repeats, which include 24 reverse repeats, 26 palindromic repeats, 34 forward repeats, and 2 complementary repeats. Among these repeats 53% (forward 22, palindrome 17, reverse 6) of the repeats are present in the IR region, 37% (reverse 18, forward 9, palindrome 3, complementary 2) in the LSC region, and 10% (palindrome 6, forward 3) in the SSC region (Supplementary Table S7). The size of reverse and complementary repeats ranges from 30 to 47, whereas the size of forward and palindrome repeats are > 100. Based on the comparison of repeats among the closest species of the genus (Park et al. 2018; Park et al. 2018; Park et al. 2018) of *E. alsinoides*, it is evident that the size of the repeats is larger (> 100). Also, the reverse and complementary repeats are very less (Park et al. 2018). Hence, it could be concluded that repetitive regions are present in higher amounts in the family Convolvulaceae than other angiosperm species.

Experimental validation of SSR markers

Out of 44 SSRs predicted by MISA, 15 representative SSRs were taken randomly for the functional validation based on GC% (> 50), primer dimer formation and hairpin secondary structure formation. The annealing temperature was fixed according to the intensity of amplified product generated with gradient PCR (Supplementary Fig. 1). Among the 15 SSR primers synthesized (Supplementary Table S8), 13 primers showed amplification with single amplified products with sizes ranging between 300 and 400 bp (Fig. 4). Based on the similarity search, 8 SSRs showed sequence homology with PCGs of chloroplast genome (Supplementary Table S9). The amplified SSRs showed a clear banding pattern in agarose gel. Among the 13 SSRs, 8 SSRs showed similarity with chloroplast PCGs such as *psbI*, *pasl*, *atpA* and *rps2*. Out of the 8 SSRs, 4 SSRs showed similarity with *psbI* (SSR3,4,5) and *pasl* (SSR6) genes, photosynthetic related genes involved in photosystem I and II. Three SSRs (SSR8,10,13) exhibited similarity with ATP synthase, and one has similarity with rps2 (SSR 12), which is involved in transcription and translation machinery. The SSR sequences were submitted to NCBI GenBank (Supplementary Table S10). Gene diversity studies could use these genes as specific markers. The SSR primers developed were highly authenticated SSR markers that could be used in genetic diversity studies in the future.

Genome comparison, sequence variation, IR contraction, and expansion

Genome comparison establishes that the LSC region is more conserved than the IR and SSC region due to gene rearrangement (Park et al. 2018; Delannoy et al. 2011; Wicke et al. 2013). Many of the protein-coding regions of the family are more conserved than the non-coding regions (Fig. 5). The IR-LSC and IR-SSC boundaries comparison of the five cp genome of the family Convolvulaceae were represented in (Fig. 5). It is evident from the results that the size of IR, SSC, and LSC regions was highly variable among the five species of the family Convolvulaceae, but the genes present in the cp genome's boundaries are not much highly variable. The genes *rpl2*, *rpl23*, *ycf1*, *ycf2*, *ycf1*, *ycf15*, *ndhH*, *ndhF*, *ndhA*, *ndhD*, *ccsA*, *rpl32*, *psbA*, *psbK, matK* are present in the IR-LSC and IR-SSC boundaries of cp genome (Fig. 6). The rpl2, rpl23, and *ycf2*

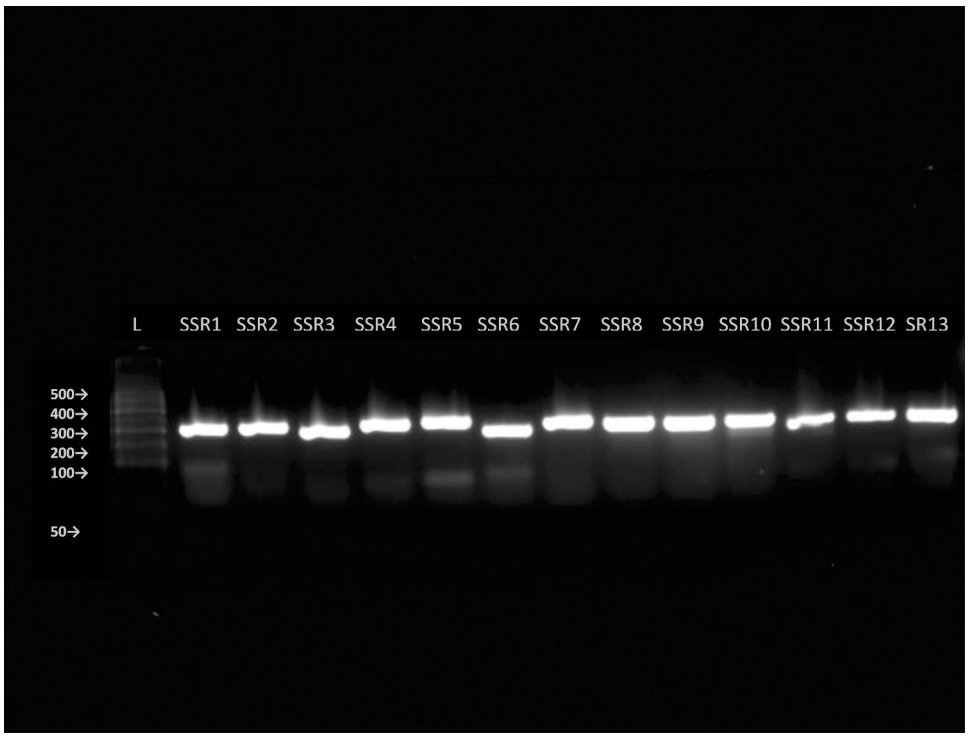

Fig. 4 Representative agarose gel image of SSR amplified regions in *E. alsinoides*. Gel image showing amplification with 3% agarose. Lane1 represents 1 kb DNA ladder and Lane2 to 14 represents SSR1 to SSR13

genes are present in the LSC-IRa boundaries. The *rpl2* and *rpl23* genes are present in all the species. *ycf2* gene was located in the LSC –IRb and LSC-IRa border in all the species. The ndhF are present in 4 species except for the genus *Cuscuta* whereas the *ndhD* gene is present in two species of *Cuscuta*. *ycf1* genes that are located in the SSC-IR boundaries are present in all the species except *I. nil.* The *ccsA* gene is present in the SSC region of all the species. *psbA* genes are present in all the five species while *E. alsinoides* has *psbK* in the IRa-LSC boundary. Many genes are common in all the species that are considered for the study. It is observed that the position and presence of few genes vary because of variation in size among the cp genomes (Huang et al. 2014; Hansen et al. 2007; Raubeson et al. 2007; Khakhlova and Bock 2006).

The size of the complete genome, the number of protein-coding genes, LSC, IR, and SSC regions are highly variable among all the Convolvulaceae species that are considered for the study. Among all the species, *Ipomoea* has the largest genome size, and *Cuscuta* has the least genome size. The LSC and SSC region was larger, and the IR region was found to be less in *C. cretica* than in all other species (Table 4). The GC content and rRNAs are almost the same in all the species, whereas the protein-coding regions and tRNAs are highly variable among all the family species. These could be the reasons for variability among genes present in LSC, IR, and SSC regions and boundaries. From the comparative analysis, it is evident that the cp genomes are highly variable among all the species of family Convolvulaceae.

Synonymous and non-synonymous substitutions rates

To estimate the selection pressure of 72 protein-coding genes in the cp genome of *E. alsinoides*, the nonsynonymous (Ka), synonymous (Ks) substitution rate and Ka/Ks ratios were calculated. For the analysis, 18 species belong to the order Solanales (seven *Ipomoea* species, *E. alsinoides*, and *C. cretica*) and nine species belong to the family Solanaceae (five Capsicum species, three *Solanum* species, and *N. tabacum*) (Supplemental Table S2). If Ka/Ks ratio > 1 the gene is influenced by positive selection, if Ka/Ks ratio < 1 the gene is influenced by negative selection, and If Ka/Ks ratio $= 1$ the gene is influenced by neutral selection (Fiz-Palacios et al. 2011). The mean Ka/Ks ratio indicates that, out of 72 protein-coding genes (Fig. 7), 54 genes were under negative selection pressure. 14 genes (*petG*, *petD*, *psbL*, *rpl16*, *atpA*, *petB*, *ndhD*, *rpl22*, *ycf3*, *ndhH*, *ccsA*, *clpP*, *ycf2*, *rpl23*) under positive selection pressure (Supplementary Table S11) and 4 genes (*psaB*, *psbE*, *petN*, *psaC*) have the Ka/Ks ratio 0. The negative selection pressure indicates that 54 genes were relatively well conserved and undergoing purifying selection. Positive selection showed that 14 genes undergo adaptive evolution in response to their environment among the species of order Solanales (Ivanova et al. 2017; Kimura

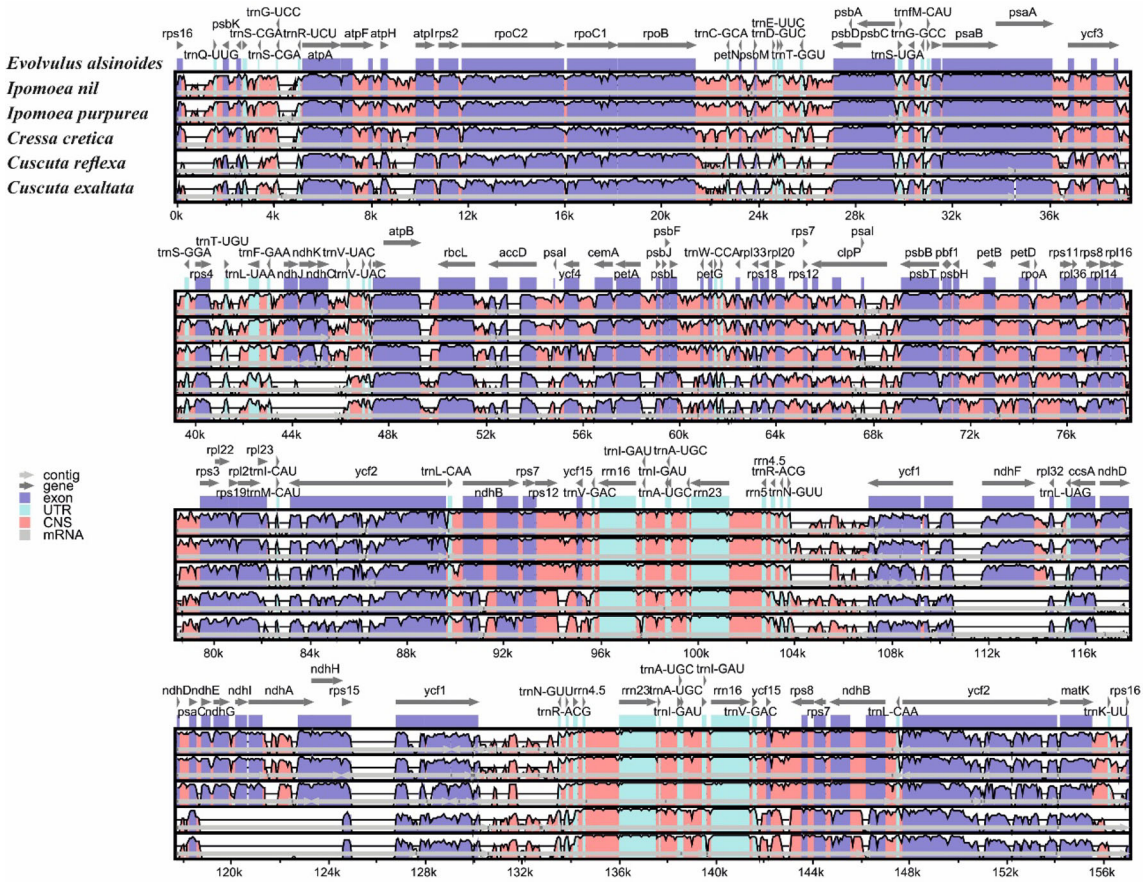

Fig. 5 Comparison of cp genomes of five species within the family Convolvulaceae with reference to the cp genome of *E. alsinoides*. The name of the genes, their position and orientation are shown above the alignment. X-axis shows the cp genome coordinates while Y-scale shows percent identity between 50 and 100%

1989; Raman and Park 2016). The Ka/Ks ratio of 0 indicates that there were no nonsynonymous (Ka) and synonymous (Ks) substitution sites in the genes, which suggests that no significant gene evolution was observed.

Divergent hotspot identification

To identify the sequence divergence level, nuclear diversity was calculated for *E. alsinoides* and species of the order Solanales (Supplementary Table S12). Sequences with Nuclear diversity values > 1.5% have been described as highly variable regions, and they can be leveraged for evolutionary studies and identification of species in seed plants (Korotkova et al. 2014; Särkinen and George 2013). It is evident from previous reports that non-coding regions were more divergent than coding regions with few exceptions of *ycf1*, *matK*, and *rbcL* genes, commonly used in plant DNA barcoding (Shaw et al. 2007; Xu et al. 2017). The nuclear diversity values ranged from 0 to 0.20, which reveals that the genomes are highly variable. Among the 18 species considered for the study, few genes such as *trnQ-UUG*, *trnF-GAA*, *PsaI*, *ClpP*, *ndhF* and *ycf1* were found to be highly variable. These divergent hot spot regions can be leveraged as phylogenetic markers for species identification, population genetics and evolutionary studies of the genus *Evolvulus* (Fig. 8). The *ndhF* gene in the cp genome was highly divergent among the species used for the present study. The ycf1 gene also has a higher divergence than other genes found in other cp genomes' coding regions (Liu et al. 2013; Nie et al. 2012; Song et al. 2015; Yukawa et al. 2006). These divergent hotspots of cp genomes can be used for species identification of the family Convolvulaceae. The IR region was less divergent than LSC and SSC regions of the cp genome (Fig. 8). The sliding window method revealed that the majority of the variations occurred in the cp genome's LSC and SSC regions (Fig. 8) (Park et al. 2018).

Phylogenetic analysis and divergence time estimation

For the present study, only Convolvulaceae and Solanaceae families are included within the order Solanales. The complete cp genome provides valuable information

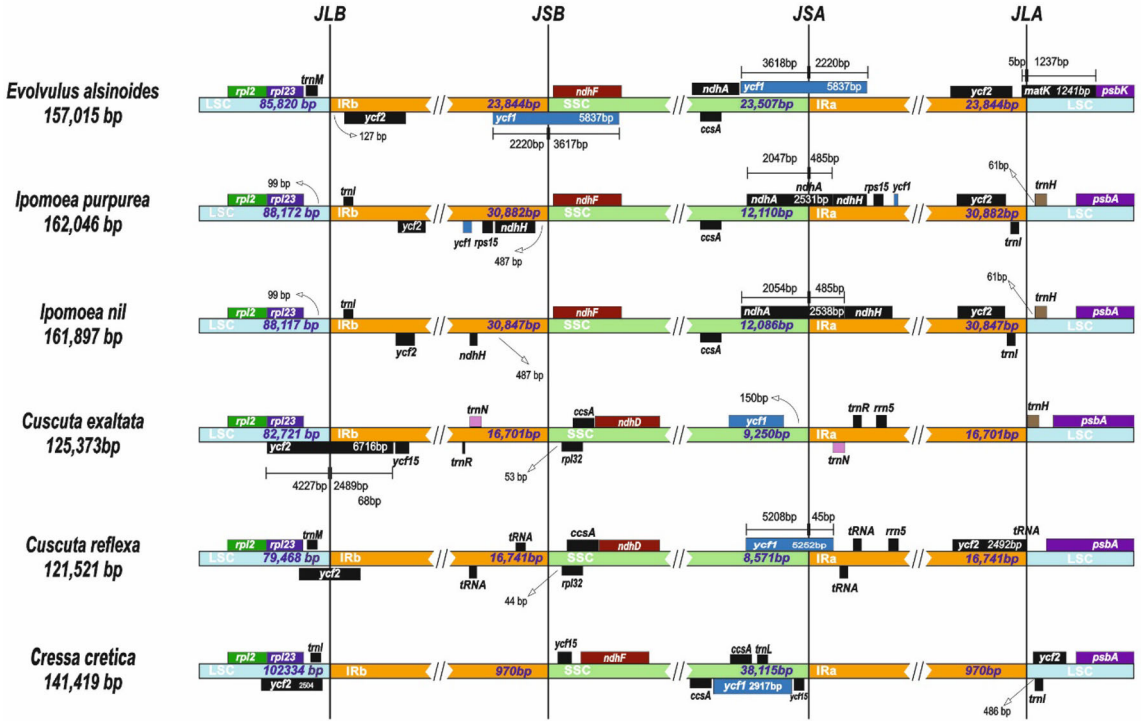

Fig. 6 Cp genome comparison of LSC, IRb, SSC and IRa border region of *E. alsinoides* with five species belongs to the family Convolvulaceae

Table 4 Comparison table showing chloroplast genome annotation of five species of the family Convolvulaceae

Species	*E. alsinoides*	*C. cretica*	*C. exaltata*	*C. reflexa*	*I. nil*	*I. purpurea*
Accession id	MN548282	MF067398	NC_009963	NC_009766	AP017304	NC_009808
Size (bp)	157,015	141,419	125,373	121,521	161,897	162,046
LSC (bp)	85,820	102,334	82,721	79,468	88,118	88,172
SSC (bp)	13,691	38,115	9250	8571	12,087	12,110
IR (bp)	57,504	970	33,402	33,482	30,846	61,764
GC%	37.3	38.6	38.1	38.2	37.5	37.5
Protein	80	78	70	61	78	78
rRNA	4	4	4	4	4	4
tRNA	30	17	19	19	30	29
Total genes	114	99	93	84	112	111

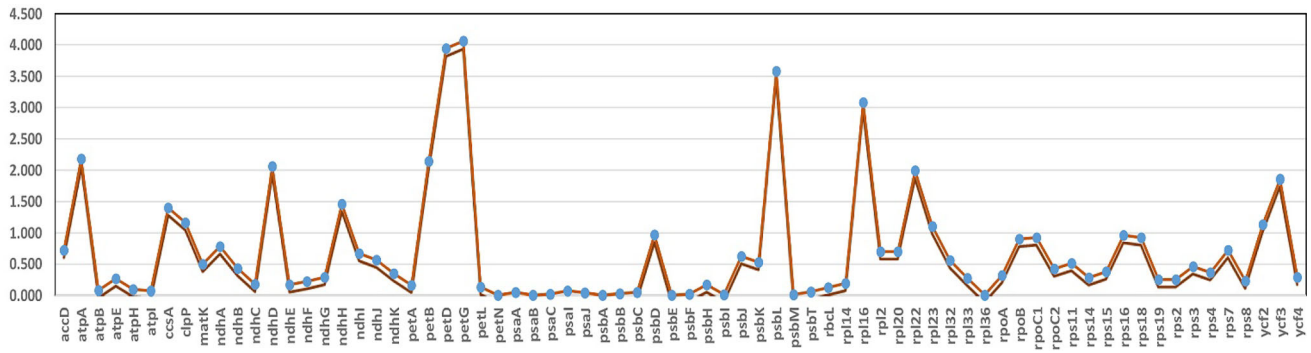

Fig. 7 Comparison of nonsynonymous (Ka) and synonymous (Ks) substitution rates among eighteen species of Solanales. Protein-coding genes name is shown in X-axis, and the mean Ka/Ks values are shown in Y-axis

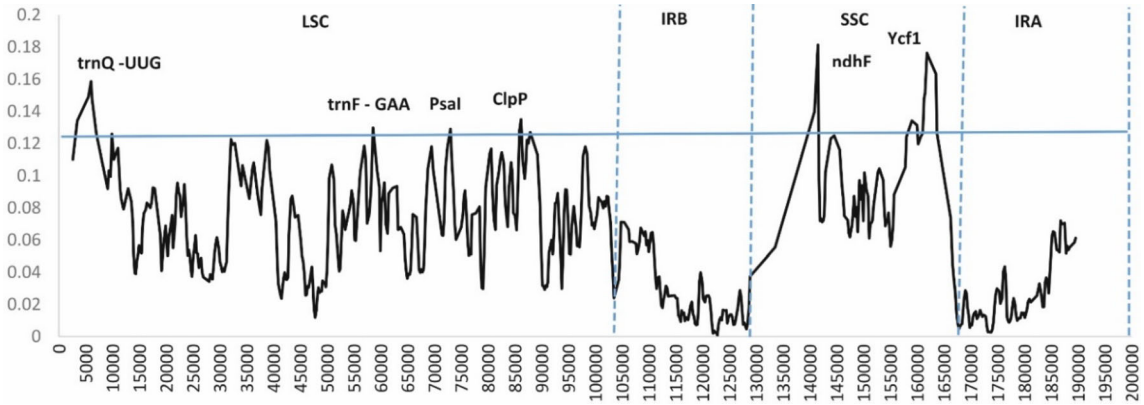

Fig. 8 Nuclear diversity (Pi) analysis of eighteen species of the order Solanales. The X-axis denotes the midpoint position of the window while the Y-axis denotes nucleotide diversity (Pi) of window. The thick blue line specifies the threshold for variation hotspots (Pi threshold = 0.13). The highly variable regions are represented in the graph

regarding the phylogenetic position of Angiosperms. Hence it could be used as an essential tool to understand the phylogeny (Austin 1978, 2008; Kim et al. 2015; Huang et al. 2014; Moore et al. 2007; Wu et al. 2010; Eserman et al. 2014). A phylogenetic tree was constructed using two datasets (1) complete cp genome and (2) concatenated alignment of 72 shared genes (Supplementary Table S13) for better understanding the phylogenetic position of *E. alsinoides*. The three methods (MP, ML, and BI) that constructed the phylogenetic tree generated the same topology tree for each dataset. The selected species formed five clades (Clade I–V), supported with 100% bootstrap (BP) values in the phylogenetic tree (Fig. 2). The clade I composed of Convolvulaceae, formed the basal position with three subclades. Subclade A included all Cuscuta species, subclade B comprised of *E. alsinoides*, and *C. cretica* and subclade C included all *Ipomoea* species. Clade II has *Petunia exserta* (Fig. 9), *Nicotiana* species formed clade III, Clade IV includes *Capsicum* species, and Clade V included all *Solanum* species. The only difference in the constructed tree using three methods (ML, MP&BI) is that the bootstrap value for a clade that belongs to the genus Ipomoea is 98% in the MP constructed tree. In ML analysis, the bootstrap value is 100 for all the clades. In BI analysis, the Bayesian posterior probabilities values are 1 for all the clades (Supplementary Fig. 2).

The divergence analysis was performed with the previously calibrated time between *I. nil* and *Arabidopsis thaliana*, and it showed they diverged 121 million years ago (Mya) (Christin et al. 2011; Fiz-Palacios et al. 2011). Our findings revealed that Clade I (Convolvulaceae) had diverged from its ancestor, Solanales, some 90 million years ago. The second speciation event happened around 68 Mya, where the family Solanaceae had evolved (Fig. 2). The Convolvulaceae species belong to Clade I and its divergence happened about 49 Mya (Fig. 2), the *Ipomoea* species diverged 39 Mya from the common ancestor of *Evolvulus* and *Cressa.* The divergence of *Evolvulus* and *Cressa* happened around 28 Mya, and the Cuscuta species divergence occurred about 15 Mya. Later, *Nicotiana* species (Clade III) diverged approximately 10 Mya (Fig. 2). Next, speciation occurred around 8 Mya (Clade IV) between the *Capsicum* and *Solanum* species, and finally, the *Solanum* species evolved about 6 Mya (Clade V). The present phylogenetic analysis showed that *E. alsinoides* and *C. cretica* form a separate sub-clade, and they diverged from the genus *Ipomoea around* 39 Mya.

Conclusion

The cp genome of *E. alsenoides* is the first completely sequenced cp genome of the genus *Evolvulus*. The cp genome had quadripartite structure with 157,015 bp in length similar to the land plant's cp genome size. The experimentally validated 15 SSR markers could be reliable species-specific markers in upcoming studies in the genus *Evolvulus*. Based on the comparison studies, 72 shared PCGs were present among the species of Solanales. Six divergent hotspots trnQ-UUG, trnF-GAA, psaI, clpP, ndhF, and ycf1 identified from the cp genome could act as molecular markers for genetic diversity studies. Fifty-four genes were conserved in the cp genome and were undergoing purifying selection. From the phylogenetic studies, it was apparent that *E. alsinoides* is closely related *C. cretica* and diverged from the genus Ipomoea about 39 Mya. The overall analysis of the cp genome shows that the size, gene content, gene orientation, and gene number of the cp genome were not highly conserved among the species of the family. The cp genome and its annotation are beneficial resources for species identification, the taxonomic position

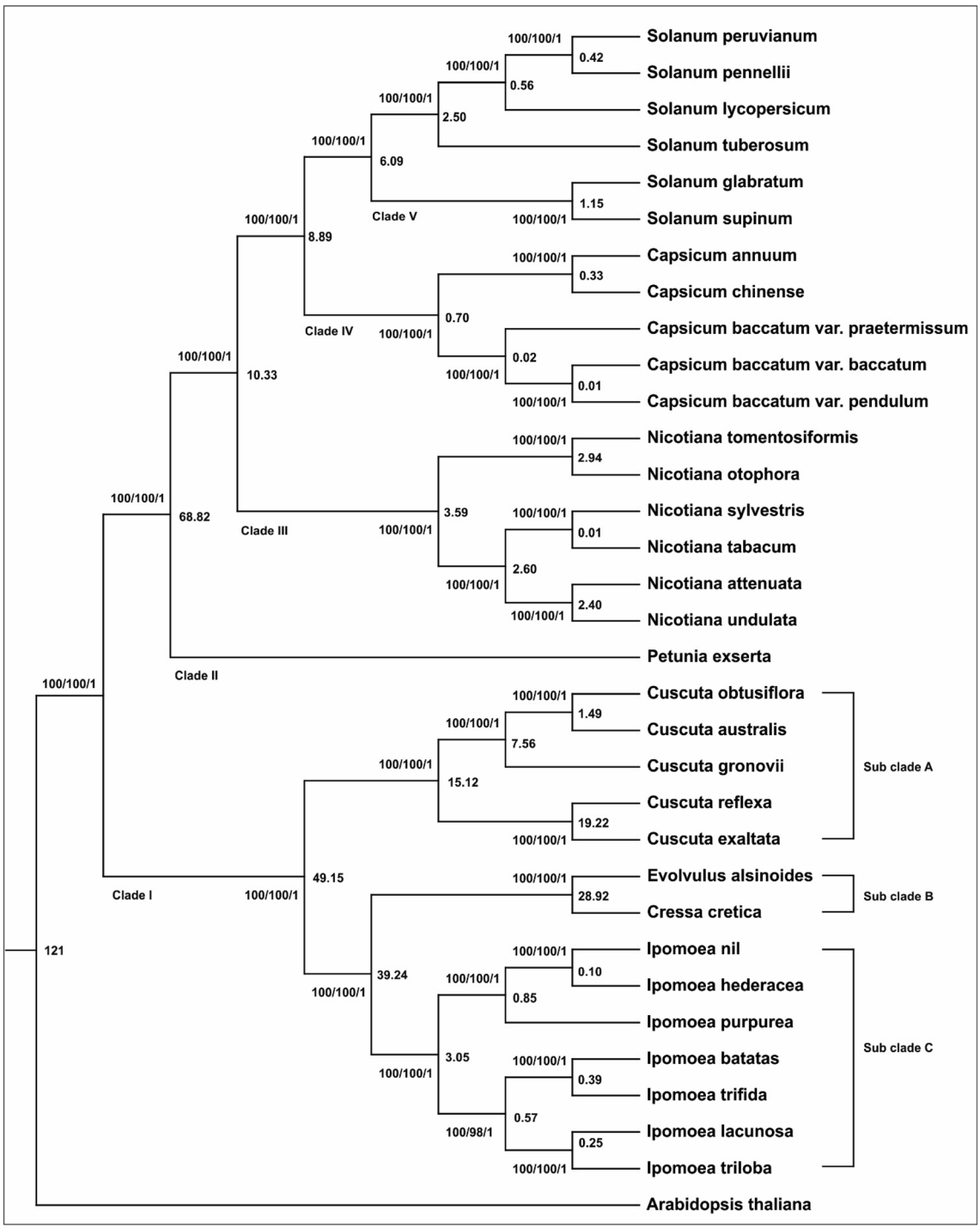

Fig. 9 Phylogenetic tree of complete cp genome constructed using BI, ML and MP methods. The values above the nodes are Bayesian posterior probabilities, MP and ML bootstrap values. The values inside the clades represent the divergence times and the divergence dates in Mya are represented in the nodes

of the species, genetic breeding, and evolutionary studies of the species of the family *Convolvulaceae*.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-021-01051-w.

Acknowledgements Our sincere gratitude to Prof. P. R. Sudhakaran and Prof. Oommen V Oommen for their valuable suggestions. The support provided by Mr. Deelip Kumar R., Campus Computing Facility (CCF), University of Kerala is gratefully acknowledged. We are grateful to the Campus Computing Facility (CCF) at the Central Laboratory for Instrumentation and Facilitation, University of Kerala for providing the HPC/GPU cluster facility to carry out this research

work. We thank AgriGenome Labs, Cochin, Kerala, India, for performing Illumina sequencing.

Author contributions SPR and ASN designed the study. SPR, VCB, SV, AS, VCL, VR, AJ and AS conducted the experiments and performed the analysis. NF conducted the experimental validation of SSR. SPR coordinated the project and wrote the manuscript.

Funding We acknowledge the funds received from the University of Kerala (Plan fund) and also the facility under SIUCEB project, DBT – BIF centre, MHRD-FAST (AiCADD center) in the Department of Computational Biology and Bioinformatics, University of Kerala.

Declarations

Conflict of interest No conflict of interest declared by authors.

References

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ambika AP, Nair SN (2019) Wound healing activity of plants from the convolvulaceae family. Adv Wound Care 8:28–37

Austin DF (1978) The Ipomoea batatas complex-I. Taxonomy. Bull Torrey Bot Club 114–129

Austin DF (2008) Author's personal copy Evolvulus alsinoides (Convolvulaceae): An American herb in the Old World. J Ethnopharmacol 117:185–198. https://doi.org/10.1016/j.jep.2008.01.038

Beier S, Thiel T, Münch T et al (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585

Biémont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. Nature 443:521–524. https://doi.org/10.1038/443521a

Bock R (2000) Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing [RNA processing, plastid transformation, evolution]. Biochim

Bryan GJ, McNicoll J, Ramsay G et al (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor Appl Genet 99:859–867

Chan PP, Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Gene prediction. Springer, pp 1–14

Chen J, Hao Z, Xu H et al (2015) The complete chloroplast genome sequence of the relict woody plant Metasequoia glyptostroboides Hu et Cheng. Front Plant Sci 6:447

Christin P-A, Osborne CP, Sage RF et al (2011) C4 eudicots are not younger than C4 monocots. J Exp Bot 62:3171–3181

De-la-Cruz IM, Núñez-Farfán J (2020) The complete chloroplast genomes of two Mexican plants of the annual herb Datura stramonium (Solanaceae). Mitochondrial DNA Part B 5:2823–2825

Delannoy E, Fujii S, Colas des Francs-Small C, et al (2011) Rampant gene loss in the underground orchid Rhizanthella gardneri highlights evolutionary constraints on plastid genomes. Mol Biol Evol 28:2077–2086

Dong W, Liu J, Yu J et al (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One 7:e35071

Dong W, Xu C, Cheng T et al (2013) Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. Genome Biol Evol 5:989–997

Du P, Jia L, Li Y (2009) CURE-Chloroplast: a chloroplast C-to-U RNA editing predictor for seed plants. BMC Bioinformatics 10:135

Duruvasula S, Mulpuri S, Kandasamy U (2019) Mapping of plastid RNA editing sites of Helianthus and identification of differential editing in fungal infected plants. Curr Plant Biol 18:100109

Ebert D, Peakall ROD (2009) 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 9:673–690

Erixon P, Oxelman B (2008) Reticulate or tree-like chloroplast DNA evolution in Sileneae (Caryophyllaceae)? Mol Phylogenet Evol 48:313–325

Ermolaeva MD (2001) Synonymous codon usage in bacteria. Curr Issues Mol Biol 3:91–97

Eserman LA, Tiley GP, Jarret RL et al (2014) Phylogenetics and diversification of morning glories (tribe Ipomoeeae, Convolvulaceae) based on whole plastome sequences. Am J Bot 101:92–103

Fiz-Palacios O, Schneider H, Heinrichs J, Savolainen V (2011) Diversification of land plants: insights from a family-level phylogenetic analysis. BMC Evol Biol 11:341

Funk HT, Berg S, Krupinska K et al (2007) Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, Cuscuta reflexa and Cuscuta gronovii. BMC Plant Biol 7:45

Hackl T, Hedrich R, Schultz J, Förster F (2014) proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics 30:3004–3011

Hansen DR, Dastidar SG, Cai Z et al (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: Buxus (Buxaceae), Chloranthus (Chloranthaceae), Dioscorea (Dioscoreaceae), and Illicium (Schisandraceae). Mol Phylogenet Evol 45:547–563

Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 10:21

Henry RJ, Rice N, Edwards M, Nock CJ (2014) Next-generation technologies to determine plastid genome sequences. In: Chloroplast biotechnology. Springer, pp 39–46

Huang H, Shi C, Liu Y et al (2014) Thirteen Camelliachloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol 14:151

Huelsenbeck JP, Ronquist F (2005) MrBayes: a program for the Bayesian inference of phylogeny, v. 3.1. 2. Rochester New York

Ivanova Z, Sablok G, Daskalova E et al (2017) Chloroplast genome analysis of resurrection tertiary relict Haberlea rhodopensis highlights genes important for desiccation stress response. Front Plant Sci 8:204

Jansen RK, Cai Z, Raubeson LA et al (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci 104:19369–19374

Jones A, Kang J (2015) Development of leaf lobing and vein pattern architecture in the Genus Ipomoea (morning glory). Int J Plant Sci 176:820–831

Kaila T, Chaduvla PK, Saxena S et al (2016) Chloroplast genome sequence of Pigeonpea (Cajanus cajan (L.) Millspaugh) and Cajanus scarabaeoides (L.) Thouars: genome organization and comparison with other legumes. Front Plant Sci 7:1847

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J 46:85–94

Kim K, Lee S-C, Lee J et al (2015) Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. Sci Rep 5:15655

Kimura M (1989) The neutral theory of molecular evolution and the world view of the neutralists. Genome 31:24–31

Koren S, Walenz BP, Berlin K et al (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736

Korotkova N, Nauheimer L, Ter-Voskanyan H et al (2014) Variability among the most rapidly evolving plastid genomic regions is lineage-specific: implications of pairwise genome comparisons in Pyrus (Rosaceae) and other angiosperms for marker choice. PLoS One 9:e112998

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 34:1812–1819

Kurtz S, Choudhuri JV, Ohlebusch E et al (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100

Li WLS, Drummond AJ (2012) Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. Mol Biol Evol 29:751–761

Liu Y, Huo N, Dong L et al (2013) Complete chloroplast genome sequences of Mongolia medicine Artemisia frigida and phylogenetic relationships with other plants. PLoS One 8:e57533

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41:W575–W581

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57

Madhavan V, Yoganarasimhan N, Gurudeva MR (2008) Pharmacognostical studies on Sankhapushpi (Convolvulus microphyllus Sieb. ex Spreng. and Evolvulus alsinoides (L.) L

Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J 17:10–12

Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci 104:19363–19368

Nie X, Lv S, Zhang Y et al (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (Ageratina adenophora). PLoS One 7:e36869

Park I, Yang S, Kim WJ et al (2018) The complete chloroplast genomes of six Ipomoea species and indel marker development for the discrimination of authentic Pharbitidis Semen (Seeds of I. nil or I. purpurea). Front Plant Sci 9:965

Park I, Yang S, Kim WJ et al (2019) Sequencing and comparative analysis of the chloroplast genome of Angelica polymorpha and the development of a novel indel marker for species identification. Molecules 24:1038

Powell W, Morgante M, McDevitt R et al (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci 92:7759–7763

Priya T (2017) Antimicrobial activity of Evovulus Alisinoids (L) extract with different organic solvents in pathogenic bacteria and fungal species

Qian J, Song J, Gao H, et al (2013) The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PLoS One 8:e57607

Qingpo L, Qingzhong X (2004) Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. *japonica*). Zuo Wu Xue Bao 30:1220–1224

Raman G, Park S (2016) The complete chloroplast genome sequence of Ampelopsis: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. Front Plant Sci 7:341

Raubeson LA, Peery R, Chumley TW et al (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. BMC Genomics 8:174

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. Plant Syst Evol 271:101–122

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Saina JK, Gichira AW, Li Z-Z et al (2018) The complete chloroplast genome sequence of Dodonaea viscosa: comparative and phylogenetic analyses. Genetica 146:101–113

Särkinen T, George M (2013) Predicting plastid marker variation: can complete plastid genomes from closely related species help? PLoS One 8:e82266

Sethiya NK, Trivedi A, Patel MB, Mishra SH (2010) Comparative pharmacognostical investigation on four ethanobotanicals traditionally used as Shankhpushpi in India. J Adv Pharm Technol Res. https://doi.org/10.4103/0110-5558.76437

Sethiya NK, Nahata A, Singh PK, Mishra SH (2019) Neuropharmacological evaluation on four traditional herbs used as nervine tonic and commonly available as Shankhpushpi in India. J Ayurveda Integr Med 10:25–31

Sharp PM, Cowe E (1991) Synonymous codon usage in Saccharomyces cerevisiae. Yeast 7:657–678

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275–288

Singh A (2008) Review of Ethnomedicinal Uses and Pharmacology of Evolvulus alsinoides Linn. Ethnobot Leafl 2008:100

Siraj MB, Khan AA, Jahangir U (2019) Therapeutic potential of Evolvulus alsinoides. J Drug Deliv Ther 9:696–701

Sloan DB, Triant DA, Forrester NJ et al (2014) A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol Phylogenet Evol 72:82–89

Song Y, Dong W, Liu B et al (2015) Comparative analysis of complete chloroplast genome sequences of two tropical trees Machilus yunnanensis and Machilus balansae in the family Lauraceae. Front Plant Sci 6:662

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Sugiura M (2005) History of chloroplast genomics. In: Discoveries in photosynthesis. Springer, pp 1057–1063

Supriya R, Priyadarshan PM (2019) Genomic technologies for Hevea breeding. In: Advances in genetics. Elsevier, pp 1–73

Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11

Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40(15):e115

Uthaman A, Nair SN (2017) A review on ten sacred flowers in Kerala: Dasapushpam. Res J Pharm Technol 10:1555–1562

Wang S, Shi C, Gao L-Z (2013) Plastid genome sequence of a wild woody oil species, Prinsepia utilis, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. PLoS One 8:e73946

Wicke S, Schneeweiss GM, Depamphilis CW et al (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76:273–297

Wicke S, Müller KF, de Pamphilis CW et al (2013) Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 25:3711–3725

Wilgenbusch JC, Swofford D (2003) Inferring evolutionary trees with PAUP. Curr Protoc Bioinforma 4–6

Wong GK-S, Wang J, Tao L et al (2002) Compositional gradients in Gramineae genes. Genome Res 12:851–856

Wu F-H, Chan M-T, Liao D-C et al (2010) Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol 10:68

Xu Q, Xiong G, Li P et al (2012) Analysis of complete nucleotide sequences of 12 Gossypium chloroplast genomes: origin and evolution of allotetraploids. PLoS One 7:e37128

Xu C, Dong W, Li W et al (2017) Comparative analysis of six Lagerstroemia complete chloroplast genomes. Front Plant Sci 8:15

Yang Y, Zhou T, Duan D et al (2016) Comparative analysis of the complete chloroplast genomes of five Quercus species. Front Plant Sci 7:959

Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol 17:1081–1090

Yukawa M, Tsudzuki T, Sugiura M (2006) The chloroplast genome of Nicotiana sylvestris and Nicotiana tomentosiformis: complete sequencing confirms that the Nicotiana sylvestris progenitor is the maternal genome donor of Nicotiana tabacum. Mol Genet Genomics 275:367–373

Zhang Y-J, Ma P-F, Li D-Z (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS One 6:e20596

Zhao C, Chen S, Sun K et al (2019) Sequencing and characterization the complete chloroplast genome of the potato, Solanum tuberosum L. Mitochondrial DNA Part B 4:953–954

Zhou T, Chen C, Wei Y et al (2016) Comparative transcriptome and chloroplast genome analyses of two related Dipteronia species. Front Plant Sci 7:1512