THE CROP JOURNAL XX (2016) XXX–XXX

HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific PCR for development of *Chenopodium quinoa*-specific markers

*Rajkumari Jashmi Devi, Nikhil K. Chrungoo**

Centre for Advanced Studies in Botany, North Eastern Hill University, Shillong, India

ARTICLE INFO

Article history:
Received 1 March 2016
Received in revised form
29 June 2016
Accepted 18 September 2016
Available online 24 September 2016

Keywords:
Chenopodium quinoa
Chenopodium album
Million years ago (Mya)
Single-nucleotide polymorphism (SNP)
Allele-specific primer extension

ABSTRACT

The genus *Chenopodium* comprises about 150 species, of which *Chenopodium quinoa* and *C. album* are important for their nutritional value. Evaluation of variation in qualitative morphological traits of plants and SNPs in chloroplast *rbc*L and *mat*K gene sequences in 19 accessions representing *C. quinoa* and *C. album* indicated that the accessions IC-411824 and IC-411825, which have white seeds, belong to *C. quinoa* rather than *C. album*. This observation was also supported by a time tree that indicated IC-411824 and IC-411825 to be a sister clade to accessions of *C. quinoa* with an estimated age of 1.2 Mya. Whereas multiple alignments of *rbc*L gene sequences from the 19 accessions revealed 1.26% parsimony-informative sites with 0.68% interspecific sequence diversity, alignment of nucleotide sequences of amplicons representing the *mat*K gene revealed 4.97% parsimony-informative sites and 2.81% interspecific sequence diversity. Validation of SNPs in the cp *rbc*L and *mat*K regions of 36 accessions belonging to *C. quinoa* and *C. album* was performed by allele-specific PCR with primers carrying a single base change at the 3′ end. We report the first *C. quinoa*-specific SNP-based primer, R1RQ-AFR, designed from *rbc*L sequences, that could differentiate *quinoa* from 64 genera including 13 species of the genus *Chenopodium*. With an estimated age of 10.5–4.1 million years (Myr), the Himalayan chenopods are evolutionarily younger than the Andean chenopods. The results establish the paraphyletic origin of the genus *Chenopodium*.

1. Introduction

Chenopodium is the second largest genus and one of the taxonomically most complex genera of the subfamily Chenopodioideae in the family Amaranthaceae [1]. Whereas the leaves of *C. album* are a source of vitamins and micronutrients [2], *C. quinoa* is important for the gluten-free flour and high protein content of its grains [3,4]. The highly polymorphic habit of species in this genus has caused many difficulties in their proper taxonomic identification. The difficulties arise mainly because of the presence of polymorphisms in the species, parallel variations between different species, occurrence of phenotypic plasticity, and presence of putative hybrids [5–8]. Whereas Wilson [5]) has described the genus *Chenopodium* as a "taxonomic receptacle", Rahiminejad and Gornall [1]) have described it as a complex group which lacked

* *Corresponding author.*
E-mail address: nchrungoo@nehu.ac.in (N.K. Chrungoo).
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

http://dx.doi.org/10.1016/j.cj.2016.06.019

good morphological characteristics to distinguish between species. Studying the genus, Wahl [9] wrote "no group of comparable size and wide distribution known to the writer has suffered the lack of understanding of the taxa involved as has the genus *Chenopodium*, especially those members of its section *Chenopodium* that are closely related to *C. album* and *C. berlandieri*." Although Wahl [9] made subsectional distinctions in the genus on the basis of variations in inflorescences and pericarps, Cole [10] subdivided the genus *Chenopodium* into four subsections on the basis of seed coat morphology.

Conventional approaches towards assessment of diversity in a species by phenotypic characterization have inherent limitations, as the information generated is often limited and expression of quantitative traits is subject to strong environmental influence. In contrast, molecular markers such as isozyme patterns, seed storage protein polymorphism, RFLP, and RAPD provide virtually unlimited information about inter- as well as intraspecific variation [11–14]. Although the low evolutionary rate of cpDNA makes it unsuitable for assessment of phylogenetic relationships among closely related taxa, Clegg and Zurawski [15] suggested that the slow evolution of the chloroplast genome makes it an ideal system for assessing plant phylogeny. Besides being one of the most rapidly evolving protein coding regions of the chloroplast genome, *mat*K exhibits an exceptional feature: equal distribution of nucleotide substitutions at the first, second, and third codon positions [17]). The gene has been shown to provide a high degree of information on phylogeny even at the deepest nodes [16–18]. Whereas Selvaraj et al. [19] have exploited the *mat*K gene to resolve family- and species-level relationships in Zingiberaceae, Bafeel et al. [20] demonstrated the ability of *mat*K and *rbc*L genes, either singly or in combination, to differentiate between species of the genus *Chenopodium*. *rbc*L and *mat*K loci have also been suggested [21] to meet, at least in part, the requirements of a DNA-based barcode for plants. Although the phylogeny of major lineages in the Chenopodiaceae is poorly understood, sequence information from cp*rbc*L [22] and *mat*K/*trn*K regions [23] indicates the subfamily Chenopodioideae to be monophyletic within the family Chenopodiaceae. Maximum parsimony and Bayesian analyses of the non-coding cp*trn*L-F and nrITS regions of *Chenopodium* sensu *lato*, however, suggests that the genus is highly paraphyletic, with five major clades within the subfamily Chenopodioideae [24].

Most work on genetic diversity and phylogeny in *Chenopodium* has focused on *C. quinoa* and *C. berlandieri* subsp. *nuttalliae* and only a few studies have involved other important species of the genus including *C. album*. Previous studies aimed at elucidating this taxonomic complex on the basis of cytology [25,26], karyotypic analysis [27,28], flavonoids [1], RAPD profiles [29,30], and ISSR markers [31] clearly indicate the status of *C. album* as the most polymorphic species of the genus *Chenopodium*. While some researchers have recognized several intergrading sub-species within *C. album*, others have developed elaborate intraspecific hierarchies with numerous subspecies, morphotypes, and/or submorphotypes. Neither approach has, however, led to unequivocal resolution of the problem.

The domesticated chenopods of the Himalayas have been classified into four cultivars on the basis of seed color [32]. Whereas three of these cultivars bearing respectively black, brown, and red seeds have been correctly assigned to the *C. album* complex, the fourth cultivar, bearing white seeds, is morphologically similar to *C. quinoa*, thereby raising a question as to its taxonomic status. In view of the economic importance of *C. quinoa*, the present paper describes the development of an effective chloroplast *rbc*L- and *mat*K-based molecular marker for distinguishing *C. quinoa* from other species of the genus *Chenopodium*.

2. Materials and methods

2.1. Plant materials

Seeds of 19 accessions of the genus *Chenopodium*, comprising 11 accessions of *C. quinoa* and 8 of *C. album* (Table 1), were procured from the National Bureau of Plant Genetic Resources, NBPGR, Shimla (India). Plants of each accession were raised to full maturity in the experimental garden of the North Eastern Hill University, Shillong.

2.2. Measurement of morphological traits

The accessions were evaluated for color, shape, and arrangement of leaves; flower color; pollen morphology; and color, shape and surface features of seeds. Data were recorded at maturity for 10 plants from each accession and subjected to cluster analysis using the average linkage method. Data were collected for two successive years with three replications for each accession.

The morphology of seed coat was determined by scanning electron microscopy after removal of the pericarp from the seeds and sputter-coating with gold–platinum using Jeol, JFC-1100 fine coat ion sputter. Scanning electron microscopy of pollen grains was performed after fixation of entire flowers with 3% glutaraldehyde for 4 h followed by washing with phosphate buffer and dehydration by passing through a series of increasing concentrations (30% to 100%) of acetone at 4 °C. The dehydrated flowers were mounted on a brass stub, opened with a needle to let pollen out of the pollen sac, and sputter-coated with gold-platinum. The processed seed and flower samples were scanned under a scanning electron microscope (JSM-6360, JEOL) to study the morphology of seed surface and pollen texture.

2.3. PCR amplification of targeted chloroplast regions

Total DNA to be used as a template for PCR was isolated from young leaves of all the accessions following the method of Murray and Thomson [33]. PCR amplification of chloroplast *rbc*L and *mat*K regions from DNA of each accession was performed with primer pairs jrF (5′-ATTATACTCCTGAGTATGA3-′)-jrR (5′-ACTCCATTTGCTAGCTTC3-′) designed in our laboratory and AF (5′-CTATATCCACTTATCTTTCAGGAGT3-′)-8R (5-′AA AGTTCTAGCACAAGAAAGTCGA3-′) [34], respectively. PCR was performed for 35 cycles with an initial hot start at 94 °C for 5 min and annealing temperatures of 54 °C for *rbc*L and 50 °C for *mat*K regions. The reaction mixture was electrophoresed on 1.2% agarose gel and the amplicons visualized under UV light in a Chemi Doc XRS+ system with Quantity One 1-D analysis software version 4.6.9 (Bio-Rad). All amplifications were performed in triplicate with five samples from each accession.

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

Table 1 – Accessions of *Chenopodium quinoa* and *C. album* investigated in the present study for their qualitative morphological characters: leaf color, leaf arrangement, leaf shape, leaf apex, leaf margin, seed color, seed texture, seed coat, seed edge, pollen aperture, and pollen type.

Sl. No.	Accession	Species	Leaf color	Leaf arrangement	Leaf shape	Leaf apex	Leaf margin	Seed color	Seed texture	Seed coat	Seed edge	Pollen texture	Pollen type
1	IC341704	*C. album*	Green	Alternate	Lanceolate	Obtuse	Entire	Black	Shiny	Reticulate	Pattern	Sunken	Perforate
2	NIC22517	*C. album*	Red	Alternate	Rhombic	Acute	Dentate	Black	Shiny	Smooth	Pattern	Sunken	Perforate
3	IC341700	*C. album*	Red	Alternate	Rhombic	Acute	Dentate	Black	Shiny	Reticulate	Pattern	Sunken	Perforate
4	IC447575	*C. album*	Red	Alternate	Rhombic	Acute	Dentate	Brown	Shiny	Smooth	Smooth	Sunken	Perforate
5	EC359447	*C. album*	Red	Alternate	Rhombic	Acute	Dentate	Black	Shiny	Smooth	Pattern	Sunken	Perforate
6	EC359451	*C. album*	Red	Alternate	Rhombic	Acute	Dentate	Black	Shiny	Smooth	Pattern	Sunken	Perforate
7	IC411824[a]	*C. album*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
8	IC411825[a]	*C. album*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
9	EC507738	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
10	EC507739	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
11	EC5077391	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
12	EC507740	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
13	EC5077401	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
14	EC5077402	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
15	EC507741	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
16	EC507742	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
17	EC507744	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate
18	EC507747	*C. quinoa*	Green	Alternate	Rhombic	Acute	Entire	White	Dull	Smooth	Smooth	Sunken	Foveate
19	EC507748	*C. quinoa*	Green	Alternate	Rhombic	Acute	Dentate	White	Dull	Smooth	Smooth	Sunken	Foveate

[a] Accessions reported as *Chenopodium album* by NBPGR, India.

2.4. Nucleotide sequencing and sequence analysis

Nucleotide sequencing of the amplicons was performed in an Applied Biosystems ABI 3130 automated DNA sequencer. The sequences were individually subjected to BLAST searches for determining similarity with other known sequences in the GenBank databases. Statistical analyses for evaluating inter- and intraspecific sequence diversity were performed using SeqState 1.21 [23]. Multiple sequence alignments were produced with MUSCLE [35] using MEGA 6.06 [21,36] to identify SNPs in the aligned sequences.

2.5. Age estimation

The data set for age estimation included 87 sequences of *rbc*L and *mat*K genes representing 62 species from 19 families of the order Caryophyllales including nine species of the genus *Chenopodium*, 4 species of the family Ranunculaceae, and one species of the family Amborellaceae (Table S1). The reported ages of the Polygonaceae, Cactaceae, Ranunculaceae, and Amborellaceae families were used as calibration nodes for estimating the ages of *C. quinoa* and *C. album*. The selection of species was based on availability of sequences for both *rbc*L as well as *mat*K genes to maintain congruence in the sample size and alignment. Sequences were aligned separately for each gene and the aligned sequences of both regions (1195 bp of *rbc*L and 809 bp of *mat*K) were concatenated into a single data set of 2004 nucleotides in length using MEGA. The basal angiosperm *Amborella trichopoda* of Amborellaceae, an early-diverging angiosperm that is sister to all extant angiosperms and evolutionarily distinct from other species [37], was included as the outgroup for greater accuracy of the estimated ages. The sequences were aligned with MEGA and the output was loaded in BEAUti 1.8.0 in nexus format for setting model parameters for BEAST. The estimated ages of 4 sets of taxa, namely i) *A. trichopoda* of Amborellaceae (158–179 Myr) [38], ii) *Hydratis canadensis, Xanthorhiza simplicissima, Ranunculus acris*, and *R. affinis* of Ranunculaceae (120 Myr) [39], iii) *Fagopyrum esculentum*, *F. urophyllum*, and *F. tataricum* of Polygonaceae (90.7–125.0 Myr) [40], and iv) *Opuntia dillenii*, *Rhipsalis flagelliformis*, and *Pereskia stenantha* of Cactaceae (65–90 Myr) [41,42] were used as calibration nodes for determining the evolutionary ages of the accessions of *Chenopodium*. The HYK substitution model with estimated base frequencies and gamma + invariant-distributed rate variation among sites, SRD06 model with 2 partitions: position (1 + 2), 3 with unlinked substitution rate parameters and unlink rate heterogeneity across codon position and the clock model with lognormal relaxed clock were used for data analysis. Whereas four calibration nodes were used as normal distribution in tree priors, the Yule prior was used to construct the tree with "ucld.mean" adjusted to a uniform prior of 10–0.000001. The XML file generated with BEAUti for 100 million generations was loaded in BEAST 1.8.0 for generating the time-measured phylogenetic tree. The maximum credibility tree was generated from BEAST 1.8.0 output using Tree Annotator 1.8.0 beast.bio.ed.ac.uk/treeannotator with a burn-in of 25%. Posterior probability values of 1.00 to 0.90, 0.89, to 0.70, and 0.69 to 0.50 indicated strong, moderate and weak clade support, respectively.

2.6. SNP genotyping using allele-specific PCR

SNP genotyping was performed by allele-specific PCR (ASPE) which works on the principle that extension of a primer can occur only when its 3′-end is a perfect complement to the allele present in the template DNA. The primers for ASPE were designed on the basis of SNPs detected in the *rbc*L and *mat*K sequences of the 19 accessions studied in the present

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

investigation. The nucleotide sequences of the primers used for ASPE are given in Table 2. Based on the consistency of SNPs in the *rbc*L and *mat*K sequences and the type of substitution, six SNPs (three each from the *rbc*L and *mat*K regions) were selected for designing primers for allele-specific PCR. For each SNP, two forward primers were designed; whereas one of the primers had a single-base mismatch specific to the template SNP of *C. quinoa*, at the 3′-end, the other primer had the substituted mismatch base specific to *C. album* at the 3′ end. Two reverse primers (one each for *rbc*L and *mat*K regions) without any base mismatch for the respective template sequences were also designed from the sequence data to allow PCR amplification. All the primer pairs were tested individually for amplification of the target DNA. To confirm specificity of the primers for the target species, PCR was performed with DNA from all the 19 accessions of *Chenopodium* studied in the present investigation and taxonomically closely related species including *Beta vulgaris* and *Amaranthus cordatus* of the family Amaranthaceae and *Fagopyrum esculentum* of the family Polygonaceae. Consistency of the base substitutions in the *rbc*L and *mat*K sequences was verified by multiple alignment of the sequences of the respective amplicons with 73 sequences of *mat*K region representing 4 genera of Amaranthaceae and 36 species of genus *Chenopodium* and with 107 sequences of *rbc*L region representing 66 genera of caryophyllales and 15 species of the genus *Chenopodium* retrieved from GenBank data bases.

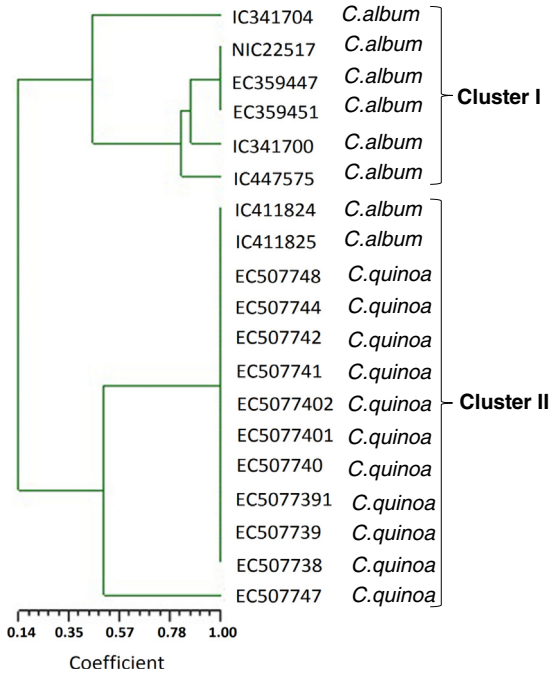

Fig. 1 – Phylogenetic tree generated from the similarity matrix developed from the scoring profile of 15 qualitative morphological characters of 19 accessions belonging to *Chenopodium quinoa* and *C. album*.

3. Results and discussion

3.1. Analysis of morphological traits

All the accessions of *Chenopodium* studied in the present investigation had yellow flowers, alternate leaf arrangement, obtuse leaf base, lenticular seeds positioned vertically in the flower, and sunken-type pollen (Table 1). Variations among the accessions were, however, observed in leaf color, leaf shape, leaf margins, type of pollen grains, color and texture of seed coat, and morphology of seed edges. Whereas five accessions of *C. album*, NIC-22517, IC-341700, IC-447575, EC-359447, and EC-359451, had reddish leaves, three

Table 2 – List of primer pairs used for allele-specific primer extension (ASPE)-based PCR genotyping of SNPs from the *rbc*L and *mat*K regions of different accessions of *Chenopodium album* and *C. quinoa* investigated in the present study.

Sl. No.	Region	SNP site	Transition/ transversion	Primer	Specific species	Sequence (5′–3′)	Reference
1	*rbc*L	C/G$_{(P'607)}$	Transversion	R1RQ[a]	*C. quinoa*	CATTACTTGAATGCTACTGC**C**	Present study
2	*rbc*L	C/G$_{(P'607)}$	Transversion	R1RC	*C. album*	CATTACTTGAATGCTACTGC**G**	Present study
3	*rbc*L	T/G$_{(P'820)}$	Transversion	R2RQ	*C. quinoa*	GCACTTCCGTGTACTAGC**T**	Present study
4	*rbc*L	T/G $_{(P'820)}$	Transversion	R2RC	*C. album*	GCACTTCCGTGTACTAGC**G**	Present study
5	*rbc*L	T/G$_{(P'1070)}$	Transversion	R3RQ	*C. quinoa*	CACGTTTGGCATATGCCTGC**T**	Present study
6	*rbc*L	T/G$_{(P'1070)}$	Transversion	R3RC	*C. album*	CACGTTTGGCATATGCCTGC**G**	Present study
7	*rbc*L	—		AFR[a]	—	GTCCCTCATTACGAGCTTGTCC	Present study
8	*mat*K	C/A$_{(P'57)}$	Transversion	R1MC	*C. album*	CGCTTTAGGGACGATCCA**C**	Present study
9	*mat*K	C/A$_{(P'57)}$	Transversion	R1MQ	*C. quinoa*	GCTTTAGCCAACGATCCA**A**	Present study
10	*mat*K	G/C$_{(P'220)}$	Transversion	R2MC	*C. album*	GATAATCGATTGATATTTT**G**	Present study
11	*mat*K	G/C$_{(P'220)}$	Transversion	R2MQ[a]	*C. quinoa*	GATAATCGATTGATATAGA**C**	Present study
12	*mat*K	G/A$_{(P'517)}$	Transition	R3MC	*C. album*	GATTCCTTTTTTTCAAAAA**G**	Present study
13	*mat*K	G/A$_{(P'517)}$	Transition	R3MQ	*C. quinoa*	GATTCCTTTTTTTCAAAAA**A**	Present study
14	*mat*K	—		AFM[a]	—	CAAACTCTTCGCTACTGGTTG	Present study

Figures in parentheses denote the position of the polymorphic base at the 3′ end in the nucleotide sequence of the respective loci. Primers AFR and AFM represent the two reverse primers for ASPE from *rbc*L and *mat*K regions, respectively. Nucleotide bases in bold at 3′ end of the primer sequences are species-specific.

[a] Indicates the most efficient *C. quinoa*-specific primer pairs for the respective target DNA segments for discriminating *C. quinoa* from *C. album*.

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

Table 3 – Gene bank accession numbers assigned to the *rbc*L and *mat*K sequences amplified from accessions of *Chenopodium quinoa* and *C. album* investigated in the present study.

Species accession number	Species	GenBank accession number	
		*rbc*L region	*mat*K region
IC-341704	*C. album*	KF319011	KF318993
NIC-22517	*C. album*	KF319012	KF318994
IC-341700	*C. album*	KF319013	KF318995
IC-447575	*C. album*	KF319014	KF318996
EC-359447	*C. album*	KF319015	KF318997
EC-359451	*C. album*	KF319016	KF318998
EC-507744	*C. album*	KF318999	KF318981
EC-507742	*C. album*	KF319000	KF318982
IC-411825[a]	*C. album*	KF319001	KF318983
IC-411824[a]	*C. album*	KF319002	KF318984
EC-507738	*C. quinoa*	KF319003	KF318985
EC-507739	*C. quinoa*	KF319004	KF318986
EC-5077391	*C. quinoa*	KF319005	KF318987
EC-507740	*C. quinoa*	KF319006	KF318988
EC-5077401	*C. quinoa*	KF319007	KF318989
EC-5077402	*C. quinoa*	KF319008	KF318990
EC-507741	*C. quinoa*	KF319009	KF318991
EC-507747	*C. quinoa*	KF709217	KF709218
EC-507748	*C. quinoa*	KF319010	KF318992

[a] Accessions reported as *Chenopodium album* by NBPGR, India.

accessions, IC-341704, IC-411824, and IC-411825, had green leaves. Except for the accession IC-341704, which had lanceolate leaves with an obtuse apex and entire margin, all the other accessions of *C. album* had rhombic-shaped leaves with an acute apex and dentate margins. All the accessions of *C. quinoa*, except EC-507747, had green, rhombic-shaped leaves with an acute apex and dentate margins. Whereas the seeds of plants belonging to accessions IC-341704, NIC-22517, IC-341700, EC-359447, and EC-359451 had a black seed coat with a shiny texture and patterned edges, those of IC-447575 had brown seed coats with a shiny texture and smooth edges. In contrast, seeds of IC-411824 and IC-411825 had a white seed coat with a dull texture and smooth edges. Except for the accessions IC-341704 and IC-341700, which had a reticulate seed coat, all other accessions belonging to *C. album* had a smooth seed coat. All accessions of *C. quinoa* had white seeds with a smooth seed coat having dull texture and smooth edge.

The dendrogram generated on the basis of Jaccard's similarity coefficient for the 15 qualitative traits clustered the accessions into two major clusters (Fig. 1). Whereas six accessions of *C. album* clustered together as one group with similarity coefficients ranging from 0.45 to 1.00, 10 accessions of *C. quinoa* and two accessions, IC-411824 and IC-411825, which have been reported as *C. album*, grouped together as one cluster with a similarity coefficient of 1.0. A distinct feature revealed by the dendrogram was rooting of the accession EC-507747 as a separate group within cluster II. The leaves of plants belonging to this accession had entire margins, unlike other accessions of this cluster, which had leaves with dentate margins. In the same context, the accession IC-341704 clustered separately within cluster I. Leaves of plants belonging to this accession were lanceolate with obtuse apex and entire margin, whereas all other accessions in this cluster had leaves with rhombic shape, acute apex, and dentate margins.

Bhargava et al. [2] reported a high degree of intraspecific variation in 16 phenotypic traits including plant height, days to flowering, days to maturity, leaf area, seed size, inflorescence

Table 4 – SNPs detected in the nucleotide sequences of *rbc*L and *mat*K genes amplified from 19 accessions of *Chenopodium*.

Accession No.	Species	SNP Position												
		rbcL				matK								
		P'_{607}[a]	P'_{820}	P'_{1042}[a]	P'_{1070}[a]	P'_{46}[a]	P'_{48}	P'_{57}[a]	P'_{220}[a]	P'_{245}	P'_{418}[a]	P'_{486}[a]	P'_{517}[a]	P'_{586}[a]
		(G/C)	(T/G)	(G/T)	(T/G)	(G/C)	(G/A)	(C/A)	(G/C)	(G/T)	(G/A)	(T/A)	(G/A)	(G/A)
IC-341704	*C. album*	G	T	G	T	G	G	C	G	G	G	T	G	G
NIC-22517	*C. album*	G	G	G	T	G	A	C	G	G	G	T	G	G
IC-341700	*C. album*	G	G	G	T	G	G	C	G	T	G	T	G	G
IC-447575	*C. album*	G	G	G	T	G	G	C	G	G	G	T	G	G
EC-359447	*C. album*	G	G	G	T	G	G	C	G	G	G	T	G	G
EC-359451	*C. album*	G	G	G	T	C	G	A	G	T	A	A	G	G
IC-411824[b]	*C. album*	C	T	T	G	C	A	A	C	T	A	A	A	A
IC-411825[b]	*C. album*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507738	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507739	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-5077391	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507740	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-5077401	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-5077402	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507741	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507742	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507744	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A
EC-507747	*C. quinoa*	C	T	T	G	C	A	A	C	T	A	A	A	A

Numbers in subscript represent the position of the bases.
[a] SNPs used for designing *C. quinoa*-specific primers with a single-base substitution at the 3′ end.
[b] Accession reported as *C. album* by NBPGR, India.

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

Table 5 – Indels, sequence statistics, and nucleotide pair frequency analysis of sequences representing chloroplast *rbc*L and *mat*K loci of accessions of *Chenopodium* investigated in the present study.

Gene locus	Taxa	Sequence statistics									Nucleotide pair frequencies			
		Length range (nt)	Mean length (nt)	Indels	% divergence	G + C (%)	CS	VS	PIS	SS	CpG (100 coverage)	ii	si	sv	R
*rbc*L	*C. album*	1180–1187	1183	17	0.832	44.30	1157 (97.47)	30 (2.53)	10 (0.84)	20 (1.68)	366	1171.00	4.00	6.00	0.57
*rbc*L	*C. quinoa*	1183–1188	1185	10	0.453	44.40	1160 (97.64)	26 (2.18)	6 (0.50)	20 (1.68)	374	1178.00	2.00	3.00	0.75
*rbc*L	*C. album* and *C. quinoa*	1180–1188	1184	22	0.682	44.40	1138 (95.79)	50 (4.21)	15 (1.26)	35 (2.95)	365	1174.00	3.00	5.00	0.57
*mat*K	*C. album*	784–793	789	18	3.619	32.7	710 (89.53)	84 (10.59)	31 (3.91)	52 (6.56)	150	758.00	11.00	18.00	0.61
*mat*K	*C. quinoa*	789–805	791	16	1.568	32.2	731 (90.81)	62 (7.70)	8 (0.99)	53 (6.58)	162	777.00	5.00	7.00	0.67
*mat*K	*C. album* and *C. quinoa*	784–805	790	32	2.811	32.5	639 (79.38)	139 (17.27)	40 (4.97)	95 (11.80)	128	766.00	8.00	14.00	0.61

CS, conserved sites; VS, variable site; PIS, parsimony-informative sites; SS, singleton sites; ii, identical pairs; si, transitional pairs; sv, transversional pairs; R, si/sv. Figures in parentheses for CS, VS, PIS, and SS represent values in percent.

length, and dry weight per plant in *C. quinoa*. Our results, however, did not reveal variations in 15 qualitative morphological traits among accessions of *C. quinoa*. The only difference was the presence of leaves with entire margins in EC-507747 instead of dentate as observed in other accessions. The only other accession having leaves with entire margins was IC-341704, which belongs to *C. album*. In contrast, *C. album* showed a high degree of morphological heterogeneity, with green/reddish leaves, black/brown/white seeds, and smooth/reticulate seed coat with smooth as well as patterned edges. Besides the black and brown seeds observed in different accessions of *C. album*, we also observed white seeds in two accessions, IC-411824 and IC-411825. Although the two accessions have been reported by the National Bureau of Plant Genetic Resources, India as *C. album*, they have foveate-type pollen grains and seeds with a dull texture and smooth surface, characters typical of *C. quinoa*. Although Karcz et al. [43] have associated tuberculate and smooth seed coats with *C. quinoa* and *C. album*, respectively, our results reveal that whereas seeds of *C. quinoa* had smooth seed coats those of *C. album* had smooth as well as reticulate seed coats. Our results thus support the presence of a high degree of heteromorphy in *C. album*.

3.2. Molecular profiling based on *rbc*L and *mat*K gene sequence variation

In order to generate molecular profiles to discriminate between *C. quinoa* and *C. album*, we amplified the chloroplast *rbc*L and *mat*K regions from 19 accessions of chenopods investigated in the present study. Nucleotide sequences of the amplicons have been deposited in NCBI GenBank and can be accessed under the assigned accession numbers (Table 3). Multiple alignment of nucleotide sequences of amplicons representing the chloroplast *rbc*L region identified 22 indels and four single-base transversions with 95.79% conserved sites, 1.26% parsimony-informative sites, and 0.68% interspecific sequence diversity with a transition/transversion ratio of 0.57 (Tables 4 and 5). In contrast, alignment of nucleotide sequences of amplicons representing the *mat*K region revealed 32 indels and nine single base changes, including five transversions and four transitions (Tables 4 and 5). The alignment revealed 17.27% conserved sites, 4.97% parsimony-informative sites, and 2.81% interspecific sequence diversity (Table 5). Codon-based alignment of the *rbc*L sequences identified all the SNPs in the sequences as synonymous. In contrast, codon-based alignment generated from *mat*K sequences identified the SNPs at P'_{46}, P'_{48}, P'_{57}, P'_{220}, P'_{245}, P'_{418}, P'_{486} as synonymous and those at P''_{516} and P'_{586} as nonsynonymous. The A▶G substitution at P'_{516} and P'_{586} caused a change in amino acid from leucine in *C. quinoa* to serine/phenylalanine in *C. album*.

3.3. Estimation of evolutionary age

The maximum credibility tree for age estimation of *Chenopodium quinoa* and *C. album* generated using four calibrated nodes for the sampled taxa is given in Fig. 2. The 19 accessions of chenopods belonging to *C. quinoa* and *C. album*, for which sequences of the *cp rbc*L and *mat*K regions were generated in the present study, resolved into two different clusters. The cluster consisting of all eleven accessions of *C. quinoa* and the accessions IC-411824 and

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

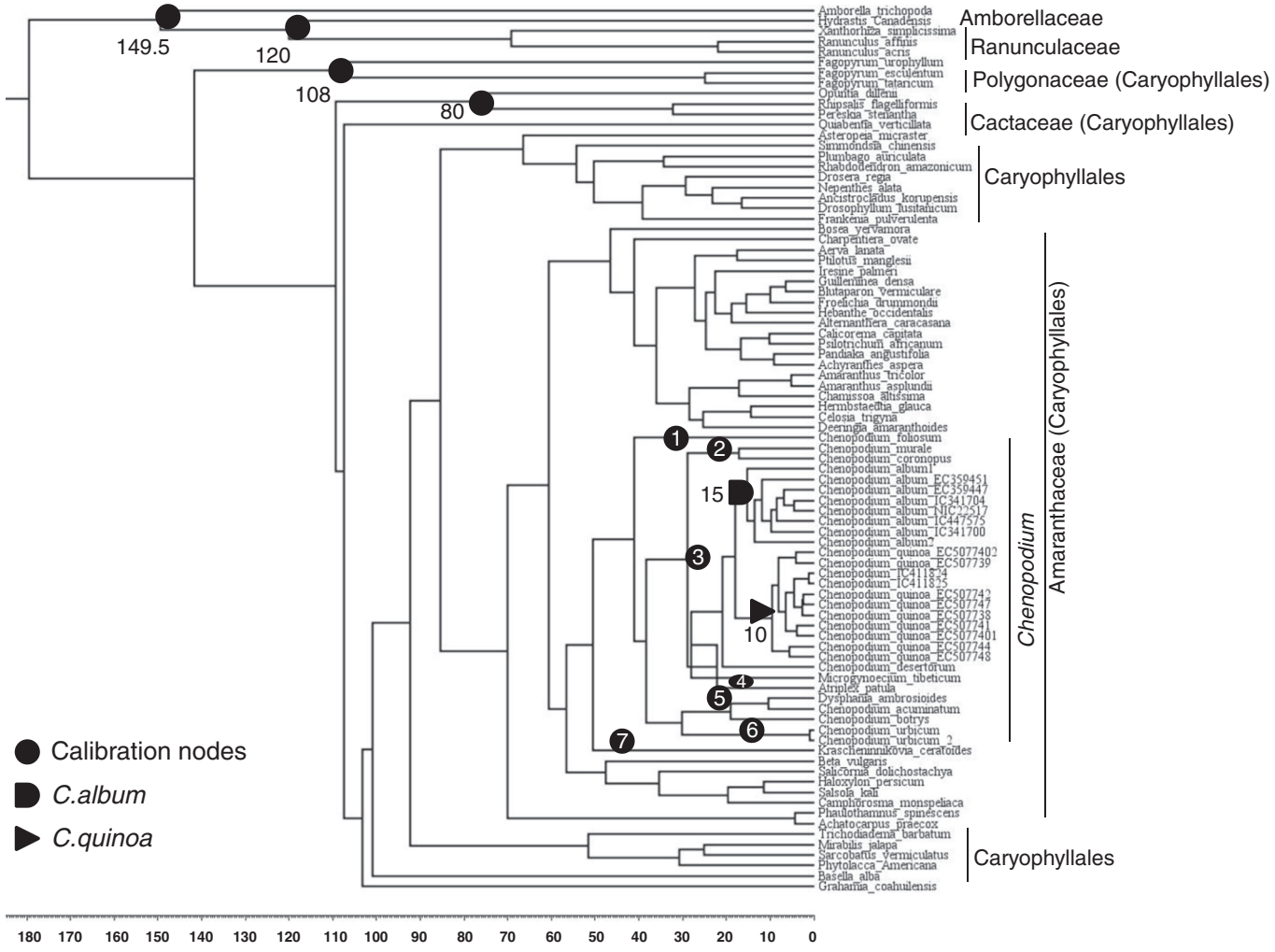

Fig. 2 – Maximum clade credibility phylogenetic tree generated with BEAST 1.8.0 from nucleotide sequence data of the *rbc*L and *mat*K regions of 46 genera belonging to 18 families of caryophyllales including 9 species of *Chenopodium*. Branch lengths in the tree correspond to age of the species. Time scale is given at the bottom of the tree in terms of millions years ago (Mya). The tree is rooted to basal angiosperm *Amoborella trichopoda* of amborellaceae (estimated as 154 Myr), whose crown age was indicated to be 179–158 Myr [38].

IC-411825 resolved as a single clade with a posterior probability value of 0.97. Whereas all the exotic accessions of *C. quinoa* were dated as having diverged 2.5–6.2 million years ago (Mya), the accessions IC-411824 and IC-411825, which have been reported by the National Bureau of Plant Genetic Resources, (India) as *C. album* but show greater similarity with *C. quinoa*, appear to have diverged 1.5 Mya. In the same context, the six accessions of *C. album* clustered together as a single clade with a posterior probability value of 0.98. Of these, four Himalayan accessions showed an estimated age of 10.5–4.1 million years (Myr) and the other two exotic accessions showed an estimated age of 13.1–7.2 Myr. These results establish a lower age for Himalayan *C. album* than for the Andean chenopods. Whereas the estimated crown age of *C. quinoa* was observed to be 13 Myr with 95% higher probability distribution estimated by the relaxed Bayesian molecular clock (HPD) ranging from 20.5–0.5, that of *C. album* was recorded as 17 Myr with 95% HPD of 22.0 to 6.0. The mean divergence time separating *C. quinoa* and *C. album* was estimated as 21 Myr. *Chenopodium foliosum* was dated as the oldest species, with an estimated age of 53.7 Myr. Whereas *C. quinoa* and *C. album* were estimated to have diverged in the Miocene epoch of the Cenozoic era (recent life), the genus *Chenopodium* appears to have originated during the Eocene epoch of the Cenozoic era. On the basis of sequence variations in the *rbc*L and ITS regions of 23 taxa of chenopodiaceae from Australia, Kadereit et al. [44] have recognized two clades in the genus *Chenopodium*. While one of the clades, comprising *C. desertorum*, *C. auricomum*, and *Rhagodia* was estimated to have an age of 4.7–2.9 Myr, the other clade, comprising *C. botrys*, *C. cristatum*, and *Dysphania*, was dated at 16.1–9.9 Myr. Our results, however, estimate the age of *Dysphania* as 30 Myr. Further, the genus formed a sister clade with *C. acuminatum*, *C. botrys*, and *C. urbicum*.

Our observations on the evolutionary lineage of different species of the genus *Chenopodium*, including other genera of the subfamily Chenopodioideae, into seven different clades are in conformity with those of Fuentes-Bazan et al. [24] and Kadereit et al. [22], thereby establishing a paraphyletic origin of the genus *Chenopodium*. Whereas Kadereit et al. [44]

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

reported the arrival of *Chenopodium* subgenera *Chenopodium/Rhagodia* in Australia during the late Miocene to Pliocene epochs when aridification and increasing salinity changed the landscape of many parts of the continent, our results indicate the origin of the genus *Chenopodium* during the Eocene to Pliocene epochs. Further, the observed divergence time of Amaranthaceae from other families of Caryophyllales is in congruence with Kadereit et al. [45] who dated the divergence of amaranthaceae to 87–47 Myr.

Although *C. quinoa* ($2n = 4\times = 36$) is reported as a tetraploid of putative allopolyploid origin [46] *C. album* is known as a complex of diploid ($2n = 18$), tetraploid ($2n = 36$) or hexaploid ($2n = 54$) species with endopolyploidy and autopolyploidy as the origin of polyploidy [28]. Although the Himalayan chenopod (*C. album*) has been suggested to be an assemblage of heteromorphic and heterocytotic (2×, 4×, 6×) forms [27,29], Joshi [47] has suggested it to be an assemblage of more than one species. Cytological investigations performed by us on accessions of *Chenopodium* studied in the present investigation identified the accessions IC-411824 and IC-411825 as tetraploid with $2n = 36$ and all other accessions belonging to *C. album* having black or brown seed as hexaploid with $2n = 56$. Given that the accessions IC-411824 and IC-411825 have been identified by National Bureau of Plant Genetic Resources, Shimla (India) as *C. album*, the variation in ploidy level indicates the cytological complexity of *C. album*. However our observations on the morphology- and

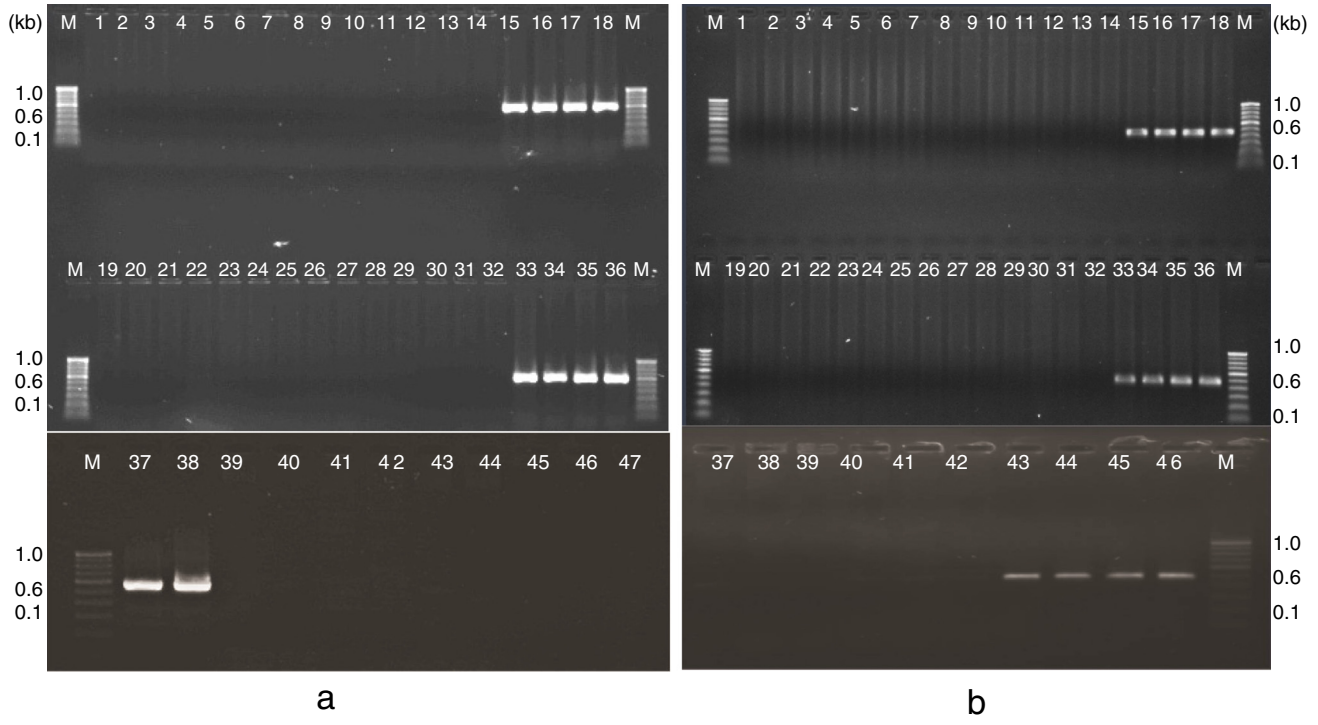

Fig. 3 – (a) Allele-specific primer based PCR amplification profiles of cp *rbc*L region with respective primer pair R1RQ-AFR from 36 accessions belonging to *Chenopodium album* and *C. quinoa*. Lanes 1: EC-349447, 2: IC-107515, 3:IC-341700, 4: IC-258332, 5: IC-313278, 6: IC-329184, 7: IC-341701, 8: IC-341707, 9: IC-341710, 10: IC-381106, 11: IC-4115421, 12: NIC-15022, 13: NIC-22489, 14: NIC-22503, 15: EC-507738, 16: EC-507739, 17: EC-507740, 18: EC-507741, 19: NIC-22504,20: NIC-22507, 21: NIC-22510, 22: NIC-22511, 23: NIC-22512,24: NIC-22513, 25: NIC-22516, 26: NIC-22518, 27: NIC-22519, 28: NIC-22520, 29: NIC-22530, 30: NIC-22533, 31: NIC-50229, 32: NIC-58617, 33: EC-507742, 34: EC-507744, 35: EC-507747, 36: EC-507748, 37: EC-507739, 38: IC-411825, 39: IC-341704, 40: NIC-22517, 41: IC-341700, 42: IC-447575, 43: EC-359447, 44: EC-359451, 45: *Fagopyrum esculentum*, 46: *Beta vulgaris*, 47: *Amaranthus cordatus*, M: 100-bp ladder. [Lanes: 1–14, 19–32, and 38–44 represent accessions belonging to *C. quinoa*, lanes: 15–18 and 33–37 represent accessions belonging to *C. album*. Lane 38 represents accession IC-411825, which has been reported as *C. album* but is suggested by us to be *C. quinoa*]. (b) Allele-specific primer based PCR amplification profiles of cp *mat*K region with primer pair R2MQ-AFM from 36 accessions belonging to *Chenopodium album* and *C. quinoa*. Lanes 1: EC-349447, 2: IC-107515, 3:IC-341700, 4: IC-258332, 5: IC-313278, 6: IC-329184, 7: IC-341701, 8: IC-341707, 9: IC-341710, 10: IC-381106, 11: IC-4115421, 12: NIC-15022, 13: NIC-22489, 14: NIC-22503, 15: EC-507738, 16: EC-507739, 17: EC-507740, 18: EC-507741, 19: NIC-22504,20: NIC-22507, 21: NIC-22510, 22: NIC-22511, 23: NIC-22512,24: NIC-22513, 25: NIC-22516, 26: NIC-22518, 27: NIC-22519, 28: NIC-22520, 29: NIC-22530, 30: NIC-22533, 31: NIC-50229, 32: NIC-58617, 33: EC-507742, 34: EC-507744, 35: EC-507747, 36: EC-507748, 37: *Amaranthus cordatus*, 38: *Beta vulgaris*, 39: *Fagopyrum esculentum* and other accessions of *Chenopodium*, namely IC-341704 (lane 40), NIC-22517 (lane 41), IC-341700 (lane 42), IC-411825 (lane 43), EC-507739 (lane 44), EC-507747 (lane 45), EC-507740 (lane 46, M: 100-bp ladder. [Lanes: 1–14, 19–32 and 40–42 represent accessions belonging to *C. quinoa*, lanes: 15–18, 33–36 and 42–46 represent accessions belonging to *C. album*. Lane 43 represents accession IC-411825, which has been reported as *C. album* but is suggested by us to be *C. quinoa*].

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

DNA-based profiles of 19 accessions of chenopods belonging to *C. quinoa* and *C. album* [48] support the greater closeness of IC-411824 and IC-411825 to *C. quinoa* than to *C. album*. In the time tree too, IC-411824 and IC-411825 formed a sister clade with accessions of *C. quinoa* with a moderate posterior probability of 0.70 and appear to have evolved during the Pleistocene epoch of the Cenozoic era (recent life) with an estimated age of 1.2 Myr.

3.4. Genotyping of SNPs by allele-specific primer extension

To develop a simple and effective molecular tool for discriminating between *C. album* and *C. quinoa*, the selected SNPs were genotyped using allele-specific primer extension (ASPE)-based PCR. For this purpose, oligonucleotide primers with a single base change at the 3′ -end, corresponding to a G/C polymorphism at P'_{607} and a G/T polymorphism at P'_{1042} and P'_{1070} in the *rbc*L gene sequence and to a G/C polymorphism at P'_{220} and a G/A polymorphism at P'_{517} and P'_{586} in the *mat*K gene sequence, were designed from the sequence data. Among the six primer pairs used for AS-PCR, the forward primer R1RQ (Table 2) with a single base mismatch for the template DNA at the 3′ -end, corresponding to the G/C polymorphism at P'_{607} in the *rbc*L sequence, in combination with the reverse primer AFR (Table 2) and the forward primer R2MQ (Table 2) with a single base mismatch for the template DNA at the 3′ -end, corresponding to the G/C polymorphism at P'_{220} in the *mat*K sequence, in combination with the reverse primer AFM (Table 2), were found to be most effective for PCR-based discrimination between *C. quinoa* and *C. album* by amplifying the target sequences only from *C. quinoa* accessions. The primer pairs amplified the target DNA from all *C. quinoa* accessions as well as IC411824 and IC411825 but did not amplify the target DNA from any other accession of *Chenopodium* or other taxonomically closely related species, including *Beta vulgaris* and *Amaranthus cordatus* of the family Amaranthaceae and *Fagopyrum esculentum* of the family Polygonaceae (Fig. 3-a, b).

Table 6 – Clustal multiple alignment of the *rbc*L and *mat*K nucleotide sequences of different species of the genus *Chenopodium* showing consistency of *C. quinoa*-specific SNP.

Sl. No.	Species	*rbc*L sequence (5′–3′)	GenBank accession number	*mat*K sequence (5′–3′)	GenBank accession number
1	*C. quinoa* (EC-5077391)	CATTACTTGAATGCTACTGC**C**	KF319005	GATAATCGATTGATATAGA**C**	KF318987
2	*C. quinoa* (EC-507747)	CATTACTTGAATGCTACTGC**C**	KF709217	GATAATCGATTGATATAGA**C**	KF709218
3	*C. quinoa* (EC-507744)	CATTACTTGAATGCTACTGC**C**	KF318999	GATAATCGATTGATATAGA**C**	KF318981
4	*C. quinoa* (EC-507742)	CATTACTTGAATGCTACTGC**C**	KF319000	GATAATCGATTGATATAGA**C**	KF318982
5	*C. album* (IC-411825)	CATTACTTGAATGCTACTGC**C**	KF319001	GATAATCGATTGATATAGA**C**	KF318983
6	*C. album* (IC-411824)	CATTACTTGAATGCTACTGC**C**	KF319002	GATAATCGATTGATATAGA**C**	KF318984
7	*C. quinoa* (EC-507738)	CATTACTTGAATGCTACTGC**C**	KF319003	GATAATCGATTGATATAGA**C**	KF318985
8	*C. quinoa* (EC-507739)	CATTACTTGAATGCTACTGC**C**	KF319004	GATAATCGATTGATATAGA**C**	KF318986
9	*C. quinoa* (EC-507740)	CATTACTTGAATGCTACTGC**C**	KF319006	GATAATCGATTGATATAGA**C**	KF318988
10	*C. quinoa* (EC-5077401)	CATTACTTGAATGCTACTGC**C**	KF319007	GATAATCGATTGATATAGA**C**	KF318989
11	*C. quinoa* (EC-5077402)	CATTACTTGAATGCTACTGC**C**	KF319008	GATAATCGATTGATATAGA**C**	KF318990
12	*C. quinoa* (EC-507741)	CATTACTTGAATGCTACTGC**C**	KF319009	GATAATCGATTGATATAGA**C**	KF318991
13	*C. quinoa* (EC-507748)	CATTACTTGAATGCTACTGC**C**	KF319010	GATAATCGATTGATATAGA**C**	KF318992
14	*C. album* (IC-341704)	CATTACTTGAATGCTACTGCG	KF319011	GATAATCGATTGATATAGAG	KF318993
15	*C. album* (NIC-22517)	CATTACTTGAATGCTACTGCG	KF319012	GATAATCGATTGATATAGAG	KF318994
16	*C. album* (IC-341700)	CATTACTTGAATGCTACTGCG	KF319013	GATAATCGATTGATATAGAG	KF318995
17	*C. album* (IC-447575)	CATTACTTGAATGCTACTGCG	KF319014	GATAATCGATTGATATAGAG	KF318996
18	*C. album* (EC-359447)	CATTACTTGAATGCTACTGCG	KF319015	GATAATCGATTGTTATAGAG	KF318997
19	*C. album* (EC-359451)	CATTACTTGAATGCTACTGCG	KF319016	GATAATCGATTGATATAGAG	KF318998
20	*C. botrys*	CATTACTTGAATGCTACTGCG	AY270080.1	GATAATCGATTGATATAGAG	AY514835.1
21	*C. frutescens*	CATTACTTGAATGCTACTGCG	AY270082.1	—	—
22	*C. cristatum*	CATTACTTGAATGCTACTGCG	AY270046.1	—	—
23	*C. auricomum*	CATTACTTGAATGCTACTGCG	AY270078.1	—	—
24	*C. desertorum*	CATTACTTGAATGCTACTGCG	AY270042.1	GATAATCGATTGATATAGAG	HE855660.1
25	*C. urbicum*	CATTACTTGAATGCTACTGCG	HM587596.1	GATAATCGATTGATATAAAG	JN895425.1
26	*C. coronopus*	CATTACTTGAATGCTACTGCG	HM587595.1	GATAATCGATTAATATAGAG	HE855636.1
27	*C. bonus-henricus*	CATTACTTGAATGCTACGGCG	AY270079.1	GATAATCGATTGATATAGAG	AF204864.1
28	*C. foliosum*	CATTACTTGAATGCTACTGCG	AY270081.1	GATAATTGATTGATATAGAG	JF953548.1
29	*C. acuminatum*	CATTACTTGAATGCTACTGCG	AY270077.1	GATAATCGATTGATATAGAG	AY514836.1
30	*C. murale*	CATTACTTGAATGCTACTGCG	HM849890.1	GATAATCGATTAATATAGAG	JQ412223.1
31	*C. glaucum*	—	—	GATAATCGATTGATATAAAG	JF953552.1
32	*C. simplex*	—	—	GATAATCGATTAATATAGAG	HQ593233.1
33	*C. capitatum*	—	—	GATAATTGATTGATATAGAG	JN966244.1
34	*C. ficifolium*	—	—	GATAATCGATTGATATAGAG	JN894385.1
35	*C. polyspermum*	—	—	GATAATCGATTGATATAGAG	JN895423.1
36	*C. rubrum*	—	—	GATAATCGATTGATATAAAG	JN895422.1
37	*C. vulvaria*	—	—	GATAATCGATTGATATAGAG	JN895424.1

— indicates that sequence information is not available, Nucleotide bases in bold and underlining at 3′ end of the sequences are *C. quinoa*-specific.

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

Alignment of the nucleotide sequences of amplicons generated in the present study with retrieved partial *rbc*L and *mat*K gene sequences of different species of *Chenopodium* confirmed the *C. quinoa*-specific G►C substitution at P'_{607} in *rbc*L and at P'_{220} in *mat*K genes (Table 6). Comparison of the identification efficiency of the sequences of amplicons derived with the two primer pairs using Geneious version 8.0.3 (http://www.geneious.com/) revealed a higher efficiency of the *rbc*L than of the *mat*K sequences in discriminating *C. quinoa* from other species of the genus *Chenopodium*. Whereas the *rbc*L sequences of *C. quinoa* and the accessions IC-411824 and IC-411825 clustered as a single group among 107 sequences representing 64 genera including 13 species of the genus *Chenopodium*, that of *C. album* clustered together with the sequence of *C. murale* (Fig. S1). Although Bafeel et al. [20] have suggested that the *matK* gene sequence had high discrimination efficiency for species identification in the genus *Chenopodium*, our results indicate that the chloroplast *rbc*L gene sequence amplified with the primer pair R1RQ-AFR has higher efficiency for discriminating between *C. quinoa* and *C. album*. Whereas protocols for AS-PCR based genotyping require two amplification reactions, we report the development of the first *C. quinoa*-specific SNP-based marker designed from the nucleotide sequence of the chloroplast *rbc*L region of the plant with a single-amplification reaction-based protocol, in which one of the primers has a single base substitution at the 3′ -end, for identification of *C. quinoa*. The results also suggest the accessions IC-411824 and IC-411825 to be *C. quinoa* rather than *C. album*.

Acknowledgments

Financial support received from Department of Biotechnology, Government of India vide grant No. BT/PR-8953/BCE/08/533/2007 and under the Biotech Hub project sanctioned against grant No. BT/04/NE/2009 is gratefully acknowledged. RJD gratefully acknowledges the receipt of financial support from Department of Science & Technology, Government of India in the form of a research fellowship under the INSPIRE program.

Appendix A. Supplementary data

Supplementary data for this article can be found online at http://dx.doi.org/10.1016/j.cj.2016.06.019.

REFERENCES

[1] M.R. Rahiminejad, R.J. Gornall, Flavonoid, evidence for allopolyploidy in the *Chenopodium album* aggregate (Amaranthaceae), Plant Syst. Evol. 246 (2004) 77–87.
[2] A. Bhargava, S. Shukla, D. Ohri, Evaluation of foliage yield and leaf quality traits in *Chenopodium* spp. in multiyear trials, Euphytica 153 (2007) 199–213.
[3] M.J. Koziol, Chemical composition and nutritional value of quinoa (*Chenopodium quinoa* Willd.), J. Food Compos. Anal. 5 (1992) 35–68.
[4] T. Partap, B.D. Joshi, N.W. Galwey, Chenopods: *Chenopodium* spp., Promoting the Conservation and Use of Underutilized and Neglected Crops 22, International Plant Genetics Research Institute, Rome, Italy, 1998.
[5] H.D. Wilson, Artificial hybridization among species of *Chenopodium* sect. *Chenopodium*, Syst. Bot. 5 (1980) 253–263.
[6] J.L. Duke, D.J. Crawford, Character compatibility and phyletic relationships in several closely related species of *Chenopodium* of the Western United States, Taxon 28 (1979) 307–314.
[7] N.S. Kurashinge, A.A. Agarwal, Phenotypic plasticity to light competition and herbivory in *Chenopodium album* (chenopodiaceae), Am. J. Bot. 92 (2005) 21–26.
[8] D.W. Zhou, T.H. Wang, I. Valentine, Phenotypic plasticity of life-history characters in response to different germination timing in two annual weeds, Can. J. Bot. 83 (2005) 28–36.
[9] H.A. Wahl, A preliminary study of the genus *Chenopodium* in North America, Bartonia 27 (1954) 1–46.
[10] M.J. Cole, Interspecific relationships and intraspecific variation of *Chenopodium album* L. In Britain. I. The taxonomic delimitation of the species, Watsonia 5 (1961) 47–58.
[11] J.A. Lai, W.C. Yang, J.Y. Hsiao, An assessment of genetic relationships in cultivated tea clones and native wild tea in Taiwan using RAPD and ISSR markers, Bot. Bull. Acad. Sin. 42 (2001) 93–100.
[12] P.J. Maughan, A. Bonifacio, E.N. Jellen, M.R. Stevens, C.E. Coleman, M. Ricks, S.L. Mason, D.E. Jarvis, B.W. Gardunia, D.J. Fairbanks, A genetic linkage map of quinoa (*Chenopodium quinoa*) based on AFLP, RAPD, and SSR markers, Theor. Appl. Genet. 109 (2004) 1188–1195.
[13] L. Anabalón-Rodríguez, M. Thomet-Isla, Comparative analysis of genetic and morphologic diversity among quinoa accessions (*Chenopodium quinoa* Willd.) of the South of Chile and highland accessions, J. Plant Breed. Crop Sci. 1 (2009) 210–216.
[14] G. Baraket, A.B. Abdelkrim, A. Salhi-Hannachi, tRNALeu intron (UAA) of *Ficus carica* L.: genetic diversity and evolutionary patterns, Genet. Mol. Res. 14 (2015) 3817–3832.
[15] M.T. Clegg, G.H. Learn, E.M. Golenberg, Molecular Evolution of Chloroplast DNA, in: R.K. Selander, A.G. Clark, T.S. Whittam (Eds.), Evolution at the Molecular Level, Sinauer Publishers, Sunderland, United Kingdom 1991, pp. 135–149.
[16] K.F. Müller, T. Borsch, K.W. Hilu, Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting matK, trnT-F, and rbcL in basal angiosperms, Mol. Phylogenet. Evol. 41 (2006) 99–117.
[17] K.W. Hilu, H. Liang, The matK gene: sequence variation and application in plant systematics, Am. J. Bot. 84 (1997) 830–839.
[18] T. Borsch, D. Quandt, Mutational dynamics and phylogenetic utility of non-coding chloroplast DNA, Plant Syst. Evol. 282 (2009) 169–199.
[19] D. Selvaraj, R.K. Sarma, R. Sathishkumar, Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding, Bioinformatics 3 (2008) 24–27.
[20] S.O. Bafeel, I.A. Arif, A.A. Al-Homaidan, H.A. Khan, A. Ahamed, M.A. Bakir, Assessment of DNA barcoding for the identification of *Chenopodium murale* L. (chenopodaaceae), Int. J. Biol. 4 (2012) 66–74.
[21] CBOL Plant Working Group, A DNA barcode for land plants, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 12794–12797, http://dx.doi.org/10.1073/pnas.0905845106.
[22] G. Kadereit, T. Borsch, K. Weising, H. Freitag, Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis, Int. J. Plant Sci. 164 (2003) 959–986.
[23] K. Müller, SeqState: primer design and sequence statistics for phylogenetic DNA datasets, Appl. Bioinformatics 4 (2005) 65–69.
[24] S. Fuentes-Bazan, G. Mansion, T. Borsch, Towards a species level tree of the globally diverse genus *Chenopodium* (Chenopodiaceae), Mol. Phylogenet. Evol. 62 (2012) 359–374.
[25] P. Mehra, C. Malik, Cytology of some Indian chenopodiaceae, Caryologia 16 (1963) 67–84.

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019

[26] K.K. Mukherjee, A comparative study of two cytotypes of *Chenopodium album* in West Bengal, India, Can. J. Bot. 64 (1986) 754–759.
[27] A. Bhargava, S. Shukla, D. Ohri, Karyotypic studies on some cultivated and wild species of *Chenopodium* (chenopodiaceae), Genet. Resour. Crop. Evol. 53 (2006) 1309–1320.
[28] B. Kolano, A. Pluciennicza k, M. Kwasniewski, J. Maluszynska, Chromosomal localization of a novel repetitive sequence in the *Chenopodium quinoa* genome, J. Appl. Genet. 49 (2008) 313–320.
[29] G. Gangopadhyay, S. Das, K.K. Mukherjee, Speciation in *Chenopodium* in west Bengal, India, Genet. Resour. Crop. Evol. 49 (2002) 503–510.
[30] T.S. Rana, D. Narzary, D. Ohri, Genetic diversity and relationships among some wild and cultivated species of *Chenopodium* L. (Amaranthaceae) using RAPD and DAMD methods, Curr. Sci. 98 (2010) 840–846.
[31] T.S. Rana, D. Narzary, D. Ohri, Molecular differentiation of *Chenopodium album* complex and some related species using ISSR profiles and ITS sequences, Gene 495 (2012) 29–35.
[32] T. Partap, P. Kapoor, The Himalayan grain chenopods ll comparative morphology, Agric. Ecosyst. Environ. 14 (1985) 201–220.
[33] M.G. Murray, W.F. Thompson, Rapid isolation of high molecular weight plant DNA, Nucleic Acids Res. 8 (1980) 4321–4325.
[34] K. Ooi, Y. Endo, J. Yokoyama, J. Murata, Useful primer designs to amplify DNA fragments of plastid gene *matK* from angiosperm plants, Jpn. J. Bot. 70 (1995) 328–331.
[35] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[36] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.
[37] L.S. Jin, A Spatio-Temporal Phylogenetic Approach to Community EcologyPh.D. thesis University of Toronto, Canada, 2015.
[38] N. Wikström, V. Savolainen, M.W. Chase, Evolution of the angiosperms: calibrating the family tree, Proc. R. Soc. Lond. B 268 (2001) 2211–2220.
[39] C.L. Anderson, K. Bremer, E.M. Friis, Dating phylogenetically basal eudicots using *rbc*L sequences and multiple fossil reference points, Am. J. Bot. 92 (2005) 1737–1748.
[40] T.M. Schuster, S.D. Setaro, K.A. Kron, Age estimates for the buckwheat family polygonaceae based on sequence data calibrated by fossils and with a focus on the Amphi-Pacific *Muehlenbeckia*, PLoS One 8 (4) (2013), e61261.
[41] A. Shmida, Biogeography of the Desert Floras of the World, in: M. Evenari, I. Noy-Meir, D.W. Goodall (Eds.), Hot Deserts Ecosystems of the World, vol. 12a, Elsevier Amsterdam, The Netherlands 1985, pp. 23–77.
[42] J.D. Mauseth, Continental drift, climate and the evolution of cacti, Cact. Succ. J. 62 (1990) 301–308.
[43] J. Karcz, B. Kolano, J. Maluszynska, SEM studies on fruit and seed of some *Chenopodium* L. species (chenopodiaceae), XII International Conference on Plant Embryology, September 5–7, 2005, Cracow, Poland, 2005.
[44] G. Kadereit, D. Gotzek, S. Jacobs, H. Freitag, Origin and age of Australian Chenopodiaceae, Org. Divers. Evol. 5 (2005) 59–80.
[45] G. Kadereit, D. Ackerly, M.D. Pirie, A broader model for C4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.), Proc. R. Soc. Lond. B 279 (1741) (2012) 3304–3311.
[46] H. Wilson, J. Manhart, Crop/weed gene flow: *Chenopodium quinoa* Willd. and *C. berlandieri* Moq, Theor. Appl. Genet. 86 (1993) 642–648.
[47] B.D. Joshi, Genetic Resources of Leaf and Grain Amaranthus and Chenopod, in: M.S. Swaminathan, S. Jana (Eds.), Biodiversity, Macmillan India Ltd., Madras, India 1991, pp. 121–134.
[48] R.J. Devi, N.K. Chrungoo, Diversity in seed and pollen morphology of *Chenopodium quinoa* and *Chenopodium album*: potential underutilized pseudo-cereal crops of genus *Chenopodium* complex, NeBIO 4 (2013) 36–41.

Please cite this article as: R.J. Devi, N.K. Chrungoo, Evolutionary divergence in *Chenopodium* and validation of SNPs in chloroplast *rbc*L and *mat*k genes by allele-specific..., The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.06.019