Hindawi
International Journal of Agronomy
Volume 2023, Article ID 6681851, 9 pages
https://doi.org/10.1155/2023/6681851

Research Article

Phylogenetic Determination of *Chenopodium quinoa* Based on the Chloroplast Genes *rbcL* and *matK*

Elsa Helena Manjarres-Hernández[1] **and Ana Cruz Morillo-Coronado**[2]

[1]*Escuela de Ciencias Biológicas, Universidad Pedagógica y Tecnológica de Colombia, Tunja 150003, Colombia*
[2]*Escuela de Ingeniería Agronómica, Universidad Pedagógica y Tecnológica de Colombia, Tunja 150003, Colombia*

Correspondence should be addressed to Elsa Helena Manjarres-Hernández; elsa.manjarres@uptc.edu.co

Received 15 July 2023; Revised 6 November 2023; Accepted 29 November 2023; Published 11 December 2023

Academic Editor: Mohamed Addi

Chenopodium quinoa is an Andean species of great interest because of its excellent nutritional quality and great adaptability to different environmental conditions. In addition, the high phenotypic diversity has caused difficulties in the correct taxonomic identification, and there are few studies on the phylogenetic relationships of quinoa in Colombia. Therefore, the objective of this research was to determine the phylogenetic relationships of quinoa with the *matK* and *rcbL* chloroplastid genes to characterize the genetic diversity in Colombian quinoa. Evolutionary analyses were performed using nucleotide substitution rates, pattern, base composition, and phylogeny construction. The *rbcL* gene presented approximately 1344 bp, and *matK* had 646 bp, which were translated into 434 and 215 amino acids, respectively. The nucleotide composition of the genes showed high percentages of similarity and identity with the *Chenopodium quinoa* sequences registered in GenBank and BOLD. Similar phylogenetic trees were obtained with the *rbcL* and *matK* genes, and both concatenated sequences grouped the accessions into clades. The results showed that Colombian quinoa has low rates of genetic differentiation that may be due to the domestication processes of the species, the lack of certified seeds, and the constant exchange of seeds between farmers in the principal producing areas of the Andean region.

1. Introduction

The cultivation of quinoa (*Chenopodium quinoa* Willd.) is vital to feeding Andean communities; however, cultivation has increased notably in regions of North America, Europe, Africa, and Asia because of its great adaptability to different environmental conditions [1]. Additionally, the quinoa grain has a high nutritional value and contains all essential amino acids, which is why it is considered a high-quality protein food global production system [2]. Studies on the morpho-agronomic characteristics of quinoa varieties in countries such as Colombia, Peru, and Bolivia have shown high phenotypic variability [1, 3–5]. However, the phenotypic characterization of quinoa varieties has limitations because information is reduced, and the expression of quantitative traits is subject to strong environmental influences [6]. Morphological parameters, such as plant height, shape and size of the leaf or stem, flower characteristics, inflorescences, and grain, are not an identification method that differentiates *Chenopodium* species or varieties since most parameters vary with age, stage of development, and environmental conditions.

An alternative for the identification of plant species and varieties is the use of molecular markers or DNA barcodes that provide more information on inter- and intraspecific variation of the species [7, 8]. Many studies have shown that DNA barcoding is an effective tool for plant identification [9–12]. Likewise, given the high conservation of its sequence, compact size, lack of recombination, and maternal inheritance, the chloroplast genome has been used for the generation of genetic markers and for phylogenetic classification [13], genetic divergence [14], and DNA barcoding systems for molecular identification [15]. Commonly studied chloroplast coding genes include *atpF-atpH*, *matK*, *rbcL*, *rpoB*, *rpoC1*, *psbK-psbL*, and *trnH-psbA* since they meet the requirements for use as DNA barcodes in plants

[16]. Hollingworth et al. (2009) recommended the combination of two loci, ribulose-1,5-bisphosphate carboxylase (*rbcL*) and maturase (*matK*), as a plant barcode based on assessments of recoverability, sequence quality, and levels [17]. For species discrimination, this two-locus barcoding can identify species and contribute to the discovery of new species [18, 19].

Studies of these two genes in *Chenopodium* showed the paraphyletic origin of this genus [6]. An accurate assessment of the magnitude and organization pattern of the genetic diversity of individuals or conserved material of Colombian quinoa is essential to understanding the genetic structure and reliably identifying each genotype, elucidating the variability of quinoa cultivars [20], which will allow adequate use of the potential of varieties along with their conservation. Therefore, understanding the phylogenetic relationships of Colombian quinoas could be an important tool in the search for unique adaptive production, quality, and resistance characteristics for breeding programs that seek to further explore the genetic background of this species [13, 21]. The objective of this study was to understand the phylogenetic relationships between Colombian quinoa with *matK* and *rcbL* markers to establish strategies in breeding and conservation programs that lead to more efficient selection processes and the identification of quinoa germplasm of agronomic, industrial, or pharmaceutical interest.

2. Materials and Methods

2.1. Plant Material. Thirty-two accessions from the Quinoa seed collection of the Biotecnología Vegetal de la Gobernación de Boyacá were evaluated (Table 1), of which nine seedlings were planted with a randomized complete block design (RCBD) under greenhouse conditions in the city of Tunja, located at an altitude of 2690 meters above sea level, with an average temperature of 13°C, relative humidity of 78%, and 12 : 12 photoperiod. Of the nine plants planted per accession, a random sample was taken for sequencing the chloroplastic genes.

2.2. DNA Extraction and PCR Amplification of Specific Regions of the Chloroplast. DNA was isolated from young leaves of each of the quinoa accessions using the modified method of Dellaporta et al. [22], visualized on 0.8% agarose gels in a Maxicell Primo EC-340 electrophoresis chamber. The concentration was determined by spectrophotometry in a Biotek EPOCH 2 device, followed by dilution using HPLC water to a total volume of 100 μl at 10 ng/μl, stored at −20°C. For the PCR amplification of the rbcL and *matK* regions of the chloroplast, the primers (5′-ATTATACTCCTGAGTATGA-3′)—jrR (5′-ACTCCATTTGCTAGCTTC-3′) and matKF (5′-CTATATCCACTTAGTATT-TTCAGGAGG′)—matKR (5′-AAAGTTCTAGCACAAGAAAGTCGA-3′) were used, respectively, according to [6]. The amplification cycles were initial temperature at 95°C for 1 min, followed by 35 cycles of denaturation at 94°C for 30 seconds. The annealing temperature was 54°C for both primers, and the extension was at 72°C for 1.5 min. The amplified products were separated on 1.2% agarose gels at 180 volts for 1.5 hours in a Maxicell Primo EC-340 electrophoresis gel system chamber and stained with ethidium bromide.

2.3. Sequencing and Sequence Analysis. Amplicon sequencing was carried out with capillary electrophoresis on an ABI PRISM 3500 XL sequencer. The editing, assembly, and alignment of the sequences were done with BioEdit v.7.0.9.0 and MEGA-X [23]. All sequence data were uploaded to the National Center for Biotechnology Information (NCBI), and the access codes of the 62 sequences are OQ148485-OQ148548 (https://www.ncbi.nlm.nih.gov/popset/?term=OQ148485). Sequences were individually subjected to BLAST searches for similarity with sequences in GenBank databases and were further evaluated on the Barcode of Life Data System v4 (BOLD) platform. Multiple alignments of the sequences were produced with MUSCLE and CLUSTAL using MEGA-X to identify SNPs in the aligned sequences. For the general characterization of the sequences, the conserved, variable, informative parsimony, and singleton sites were evaluated, the nucleotide composition of the sequences by accession was estimated, and the pattern and nucleotide substitution rates of the *matK* and *rbcL* regions were determined. To evaluate the evolutionary distances, the model with the best fit was determined, and three phylogenetic trees were generated with the selected model, the first using the *matK* region, the second using the *rbcL* region, and the third concatenating the two *matK* and *rbcL* regions with the neighbor-joining tree algorithm, the Jukes–Cantor sequence distance estimation model, and the bootstrap of 1000 replicates.

3. Results

The *rbcL* and *matK* chloroplast regions of 32 *Chenopodium* accessions from central-eastern Colombia were amplified to generate molecular profiles that discriminated between *C. quinoa* and other *Chenopodium*. The nucleotide sequences of the amplicons were deposited in the NCBI GenBank and can be accessed under the assigned accession numbers (Table 1). Multiple alignments of the nucleotide sequences representing the *rbcL* and *matK* regions of the chloroplast identified the size of the chloroplast *matK* gene at ~650 bp and the size of the *rbcL* gene at ~1350 bp. The 32 *matK* gene sequences showed that each sequence had an average of 646 bp, with 56.96% variability, with a total of 260 conserved sites, 77 informative sites by parsimony, and a transition/transversion ratio of 0.69. The *rbcL* gene sequences, despite the fact that they presented a low percentage of variability of 32.88%, had a higher PIS (parsimony-informative site) than *matK* with 85 sites and a transition/transversion ratio of 0.82 (Table 2). A total of 434 amino acids were translated from the *rbcL* gene sequences, while a total of 215 amino acids were translated with *matK*.

The sequences of the *rbcL* and *matK* genes of the Colombian quinoas evaluated in this study presented different nucleotide compositions (Table S1). The nucleotide

Table 1: *Chenopodium* accessions are molecularly characterized by the *matK* and *rbcL* genes and GenBank access number.

Sample code	Accessions	Accession number for *matK* gene	Accession number for *rbcL* gene
Q1-9	*Quinoa real*	OQ148485	OQ148517
Q2-7	*Quinoa aurora*	OQ148486	OQ148518
Q3-3	*Quinoa ceniza 1*	OQ148487	OQ148519
Q3-8	*Quinoa ceniza 2*	OQ148488	OQ148520
Q4-6	*Quinoa beteitiva*	OQ148489	OQ148521
Q5-3	*Quinoa sotaquirá*	OQ148490	OQ148522
Q6-2	*Quinoa negra 1*	OQ148491	OQ148523
Q6-7	*Quinoa negra 2*	OQ148492	OQ148524
Q7-4	*Tunkahuan ICA*	OQ148493	OQ148525
Q8-3	*Blanca de Jericó Tuta*	OQ148494	OQ148526
Q9-2	*Amarilla de maranganí*	OQ148495	OQ148527
Q10-6	*Quinoa colorado*	OQ148496	OQ148528
Q11-4	*Blanca dulce de Soracá*	OQ148497	OQ148529
Q12-3	*Piartal Chocontá*	OQ148498	OQ148530
Q13-4	*Quinoa dulce de Tuta*	OQ148499	OQ148531
Q14-3	*Quinoa semiamarga*	OQ148500	OQ148532
Q15-3	*Quinoa peruna*	OQ148501	OQ148533
Q16-4	*Quinoa siachoque*	OQ148502	OQ148534
Q17-3	*Blanca de Jericó Tuta 2*	OQ148503	OQ148535
Q18-3	*Piartal Tibasosa*	OQ148504	OQ148536
Q19-5	*Blanca de Jericó Tunja*	OQ148505	OQ148537
Q20-3	*Blanca de Jericó Toca*	OQ148506	OQ148538
Q21-8	*Cremosa malvinas*	OQ148507	OQ148539
Q22-5	*Tunkahuan Tibasosa*	OQ148508	OQ148539
Q23-3	*Tunkahuan siachoque*	OQ148509	OQ148541
Q24-6	*Mezcla Siachoque 1*	OQ148510	OQ148542
Q25-3	*Mezcla Siachoque 2*	OQ148511	OQ148543
Q26-3	*Mezcla Siachoque 3*	OQ148512	OQ148544
Q27-7	*Quinoa primavera*	OQ148513	OQ148545
Q28-1	*Quinoa galindo*	OQ148514	OQ148546
Q29-3	*Col-quinoa*	OQ148515	OQ148547
Q30-8	*Susunaga*	OQ148516	OQ148548

Table 2: Sequence statistics and nucleotide pair frequency analysis of sequences representing chloroplast *rbcL*, and *matK* loci of accessions of *Chenopodium* investigated in the present study.

Gene	Sequence statistics					Nucleotide pair frequencies			
	Mean length (pb)	CS	VS (%)	PIS	SS	ii	si	sv	R
matK	646	260	56.96	77	288	519	13	18	0.69
rbcL	1344	800	32.88	85	352	860	19	23	0.82
matK + *rbcL*	2226	453	72.46	402	1175	1360	52	81	0.65

CS: conserved sites; VS (%): variable site percentage; PIS: parsimony-informative sites; SS: singleton sites; ii: identical pairs; si: transitional pairs; sv: transversional pairs; R: si/sv.

composition showed that the *matK* gene, on average, consists of 36.0% thymine (T), 18.3% cytosine (C), 29.7% adenine (A), and 16.0% guanine (G) while the nucleotide composition of *rbcL* was on average 29.0% thymine (T), 19.8% cytosine (C), 27.0% adenine (A), and 24.2% guanine (G).

The accessions for the *matK* gene showed varied compositions, and the Q19-5 sample had different compositions than the other accessions with 29.7% thymine (T), 15.7% cytosine (C), 36.8% adenine (A), and 17.8% guanine (G). When the sequences obtained from the *rbcL* gene were compared, the composition of the nucleotides was similar, with percentages close to the average for the genes. Only accession Q9-2 had a sequence size of 518 bp, which represented approximately half the base pairs of the average sequence. The %GC was higher in *rbcL* than in *matK*, with 42.3%.

To find the best model of the evolutionary distances and phylogeny based on the two genes, the models with lower BIC (Bayesian information criterion) scores were considered the best description of the nucleotide substitution pattern during evolution, according to [22]. The analyses of the 32

nucleotide sequences showed that the best evolutionary model for the *matK* gene sequences was 3-parameter Tamura (T92 + G + I), with 65 parameters and the lowest BIC value, 5973.62. For the *rbcL* gene, the best-fit model was a 3-parameter Tamura (T92), with 63 parameters and a BIC value of 7875.96. For the concatenated *matK* and *rbcL* sequences, the 3-parameter Tamura (T92 + G + I) model had 65 parameters and a BIC value of 25590.22.

According to the Tamura 3-parameter model and *matK* gene sequences, the nucleotide substitution rate ranged from 0.05 (AC, TG, CG, and GC) to 0.14 (CT and GA). For the *rbcL* gene, the substitution rates ranged between 0.06 (AC, TG, CG, and GC) and 0.14 (CT and GA), and, for the *matK* and *rbcL* concatenated genes, it was between 0.06 (AC, TG, CG, and GC) and 0.13. (CT and GA) (Figure 1).

The nucleotide substitution pattern was evaluated at complete codon positions (1st + 2nd + 3rd nucleotide) and according to the best-fit model of the *matK* and *rbcL* sequences, which is shown in Table 3. The transitional substitution was greater than the transversional substitution in the *matK* and *rbcL* regions. However, the *matK* region exhibited a higher substitution rate from C to T. In contrast, the frequency of change from G to A, T to C, and A to G in the *rbcL* region was higher than that of *matK*.

The concatenated sequences of *matK* and *rbcL* were analyzed using BLAST to search for homology with the sequences reported in GenBank, where 96% of the sequences presented 99% identity with the species *Chenopodium quinoa* Willd. except for accession Q9-2, which presented 98.37% identity with *Chenopodium desiccatum*; however, this same accession presented 98.37% similarity with *Chenopodium atrovirens* on the BOLD platform. The BOLD results were different in three accessions, and the species returned from this database did not match those from BLAST. The Q8-3 accession showed 99.57% similarity with *Chenopodium suecicum* in BOLD, while, in BLAST, there was 99.83% identity with *Chenopodium quinoa* Willd. Accession Q23-3 presented 96.16% similarity with *Chenopodium desiccatum* in BOLD (Table 4).

The results of the sequence analyses were consistent with the phylogenetic reconstruction from the *matK* and *rbcL* sequences, which discriminated accessions based on the number of base substitutions per site, as evidenced by the tree topologies of the sequences obtained with the neighbor-joining method (Figures 2(a) and 2(b)), which showed that some accessions had high genetic variability and a low index of genetic divergence.

The phylogenetic trees obtained with the sequences of the two genes analyzed independently showed similar groupings in some accessions in the two trees; for example, the grouping for accessions Q30-8 and Q4-6 did not correspond to the place of origin.

Finally, the phylogenetic tree obtained by concatenating the *matK* and *rbcL* regions conformed to the tree built by the *matK* segment; that is, there was little variation in the conformation of the groups, except for slight changes in the branch supports (Figure 3). Comparing these results with those obtained showed similarity and coherence in the distribution of the groups.

4. Discussion

Genetic variation is important for the survival of plants since it leads to adaptation and evolution [23]. Likewise, studies on genetic diversity among cultivated plant species are essential for conservation and improvement programs [24]. This study, where 32 *Chenopodium* accessions were analyzed by evaluating *matK* and *rbcL* genes, showed that these regions contain higher amounts of adenine + thymine (A + T) than guanine + cytosine (G + C) and that nucleotide substitution rates were high, as were genetic distances, demonstrating that the use of chloroplast molecular markers plays an important role in elucidating evolutionary relationships and identifying species, such as quinoa [25].

Therefore, it is essential to understand the evolutionary causes of nucleotide changes and their patterns of variation between species since this generates genetic divergence [26]. In this study, the *matK* region exhibited a higher nucleotide substitution rate than the *rbcL* region. Ho et al. showed that the *rbcL* gene has fewer variations than other chloroplast genes [11]. In addition, this gene presents more universality in the primers than in the *matK* gene. The latter has combinations of several primers to achieve amplification of the region [27, 28]. However, the universality of primers is an important factor in the identification of unknown species because primers are more effective if they can be used to amplify different types of plants [29]. In this study, the sequencing of both *matK* and *rbcL* genes showed effectiveness in the identification of species.

The analysis of the nucleotide composition of the sequences showed that *matK* was the region with the greatest variability or polymorphic sites with respect to the *rbcL* gene. The latter region evolves slowly and is, therefore, used more frequently in phylogenetic analyses [30]. In addition, the *rbcL* gene is a more universal region and is less efficient in separating closely related taxa [16], although some studies on angiosperm species have shown that this region presents enough variation to distinguish species [31]. Meanwhile, *matK*, which presents a faster evolution, is widely used for phylogenetic analyses. In this study, the analysis of the two regions and the analysis of the translation revealed that the nucleotide composition was similar to *Chenopodium quinoa*.

Searching the BOLD and GenBank databases of the concatenated *matK* and *rbcL* genes showed 98 to 100% agreement with *Chenopodium quinoa* sequences; however, some sequences coincided with *C. desiccatum*, *C. atrovirens*, and *C. suecicum*, and this coincides with what was obtained by [32] using Bayesian and maximum parsimony analysis of the noncoding trnL-F (cpDNA) and nuclear ITS regions, where tetraploid species of the genus *Chenopodium* such as *C. berlandieri* subsp. *nuttalliae* and *C. quinoa* belong to *Chenopodium* sensu stricto. Therefore, the DNA barcoding used in this study was able to distinguish intraspecific and interspecific divergences because, among other things, of the fact that both regions have high amplification and sequencing success rates [33].

Studies on *C. murale* obtained similar results using the *matK* and *rbcL* genes, showing 100% match with *C. murale* using BLAST, while, in BOLD, the *rbcL* gene showed a high

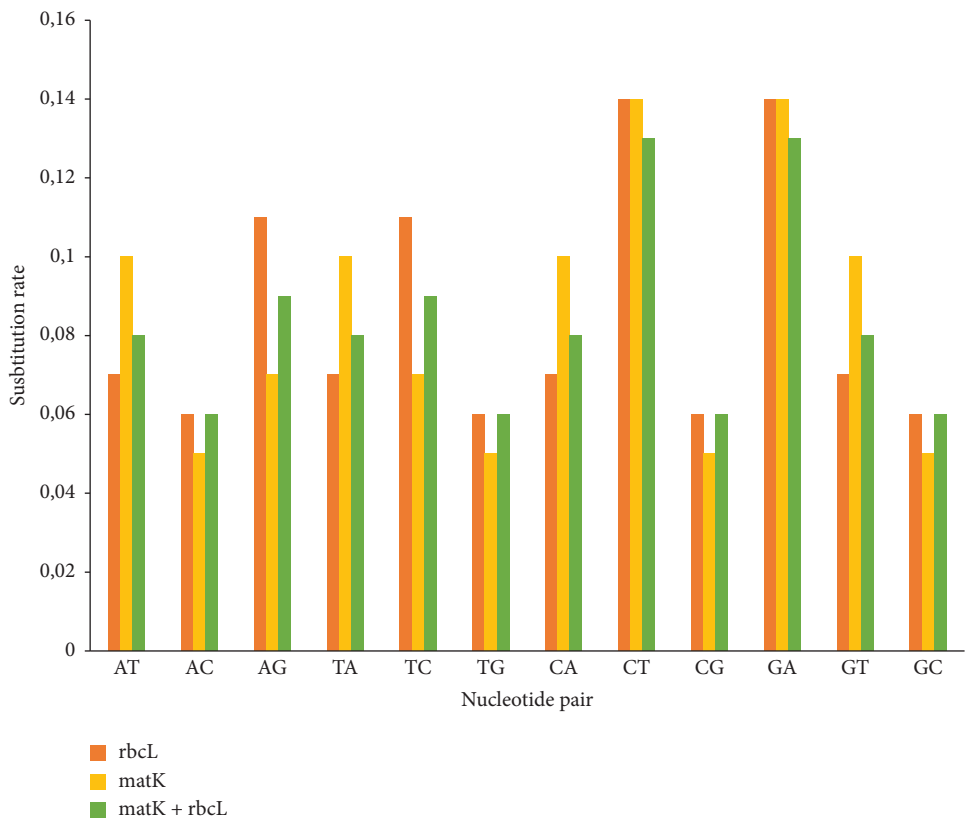

Figure 1: Nucleotide substitution rates (*matK*, *rbcL*, and *matK* + *rbcL*). The *X*-axis shows the substitution of one nucleotide for another during evolution. For example, AT implies a mutation in which adenine was replaced by thymine, and AC adenine was replaced by cytosine. The *Y*-axis shows the rate of substitution in the analyzed sequences.

Table 3: Pattern of nucleotide substitution of *matK* and *rbcL* regions (in percentage).

	A	T	C	G
matK				
A	—	*10.02*	*5.09*	**7.19**
T	*8.27*	—	**8.01**	*4.46*
C	*8.27*	**15.77**	—	*4.46*
G	**13.34**	*10.02*	*5.09*	—
rbcL				
A	—	*7.30*	*4.99*	**13.00**
T	*6.80*	—	**8.98**	*6.10*
C	*6.80*	**13.15**	—	*6.10*
G	**14.49**	*7.30*	*4.99*	—

Substitution patterns and rates were estimated with the Tamura–Nei (1993) model. Rates of different transitional substitutions are shown in bold, and the transversional substitutions are in italics.

degree of similarity with several taxa, including *C. ambrosioides*, *C. album*, and *C. ficifolium*, ranging from 96.3 to 100% [11]. Research by Huang et al. [34] showed that the combination of the *rbcL* and *matK* genes separated 40% of the sampled species in the combined data set for the family Apiaceae. Therefore, these two regions, *matK* and *rbcL*, have been shown to be ideal for species identification.

Quinoa is currently gaining international attention given the productive and nutritional potential of its seeds, which has caused a significant increase in planting areas. Because it is characterized by its primary and secondary germplasm, the authors of [14] reported the complete mitochondrial and chloroplast genome sequences of quinoa accession PI614886 and the identification of sequence variants in additional quinoa accessions and related species. This was the first reported mitochondrial genome assembly in the genus *Chenopodium*. The inference of phylogenetic relationships between *Chenopodium* species based on mitochondrial and chloroplast variants supports the hypothesis that the ancestor of the A genome was the cytoplasmic donor in the original tetraploidization event and that highland and coastal quinoas were generated independently, producing results such as those found in this study.

Ibrahim et al. 2019 characterized seven quinoa genotypes with different provenances and identified genetic polymorphisms and unique markers in each genotype with variation at the seed color level using ISSR, SCoT markers, and two chloroplastid markers (*rbcL* and *rpoC1*) [15]. Sequence alignment revealed that rbcL recovered from the quinoa genotypes with high similarity to other *rbcL* genes obtained from *Chenopodium* species in other studies, with similarities ranging between 76 and 80%. Furthermore, the *rbcL* gene showed highly consistent genetic similarity with low genetic evolution and mutation. On the other hand, the sequence alignment revealed that the *rpoC1* obtained from the quinoa genotypes had high similarity with other *rpoC1* genes obtained from other plant species in other studies. The

Table 4: Searching results of *matK* and *rbcL* genes on GenBank and BOLD databases.

Sample	BLAST with *matK* and *rbcL*	Identity (%)	BOLD with *matK* and *rbcL*	Similarity (%)
Q1-9	*Chenopodium quinoa* Willd.	99.74	*Chenopodium quinoa*	99.57
Q2-7	*Chenopodium quinoa* Willd.	100.00	*Chenopodium quinoa*	100.00
Q3-3	*Chenopodium quinoa* Willd.	99.66	*Chenopodium quinoa*	99.66
Q3-8	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.66
Q4-6	*Chenopodium quinoa* Willd.	99.75	*Chenopodium quinoa*	99.50
Q5-3	*Chenopodium quinoa* Willd.	99.57	*Chenopodium quinoa*	99.49
Q6-2	*Chenopodium quinoa* Willd.	99.66	*Chenopodium quinoa*	99.57
Q6-7	*Chenopodium quinoa* Willd.	99.92	*Chenopodium quinoa*	99.58
Q7-4	*Chenopodium quinoa* Willd.	99.92	*Chenopodium quinoa*	99.75
Q8-3	**Chenopodium quinoa Willd.**	**99.83**	**Chenopodium suecicum**	**99.57**
Q9-2	**Chenopodium desiccatum**	**98.37**	**Chenopodium atrovirens**	**98.37**
Q10-6	*Chenopodium quinoa* Willd.	99.50	*Chenopodium quinoa*	99.24
Q11-4	*Chenopodium quinoa* Willd.	98.55	*Chenopodium quinoa*	98.55
Q12-3	*Chenopodium quinoa* Willd.	99.58	*Chenopodium quinoa*	99.32
Q13-4	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.58
Q14-3	*Chenopodium quinoa* Willd.	99.75	*Chenopodium quinoa*	99.49
Q15-3	*Chenopodium quinoa* Willd.	99.75	*Chenopodium quinoa*	99.58
Q16-4	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.74
Q17-3	*Chenopodium quinoa* Willd.	99.74	*Chenopodium quinoa*	99.57
Q18-3	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.58
Q19-5	*Chenopodium quinoa* Willd	99.92	*Chenopodium quinoa*	99.75
Q20-3	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.75
Q21-8	*Chenopodium quinoa* Willd.	98.08	*Chenopodium quinoa*	97.91
Q22-5	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.58
Q23-3	**Chenopodium quinoa Willd.**	**96.96**	**Chenopodium desiccatum**	**96.16**
Q24-6	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.58
Q25-3	*Chenopodium quinoa* Willd.	100.00	*Chenopodium quinoa*	99.83
Q26-3	*Chenopodium quinoa* Willd.	95.66	*Chenopodium quinoa*	95.66
Q27-7	*Chenopodium quinoa* Willd.	99.70	*Chenopodium quinoa*	99.70
Q28-1	*Chenopodium quinoa* Willd.	99.83	*Chenopodium quinoa*	99.66
Q29-3	*Chenopodium quinoa* Willd.	99.74	*Chenopodium quinoa*	99.49
Q30-8	*Chenopodium quinoa* Willd.	99.50	*Chenopodium quinoa*	99.32
Mean		99.42		99.24

*Bold values indicate that there is no match between GenBank and BOLD databases.

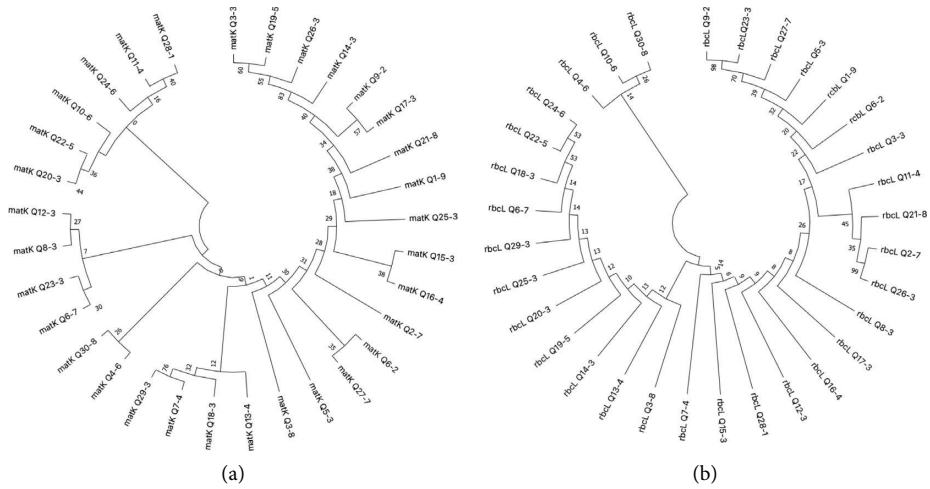

Figure 2: Phylogenetic tree based on the sequences of 32 quinoa accessions by neighbor-joining method. (a) Sequences of *matK*. (b) Sequences of *rbcL*.

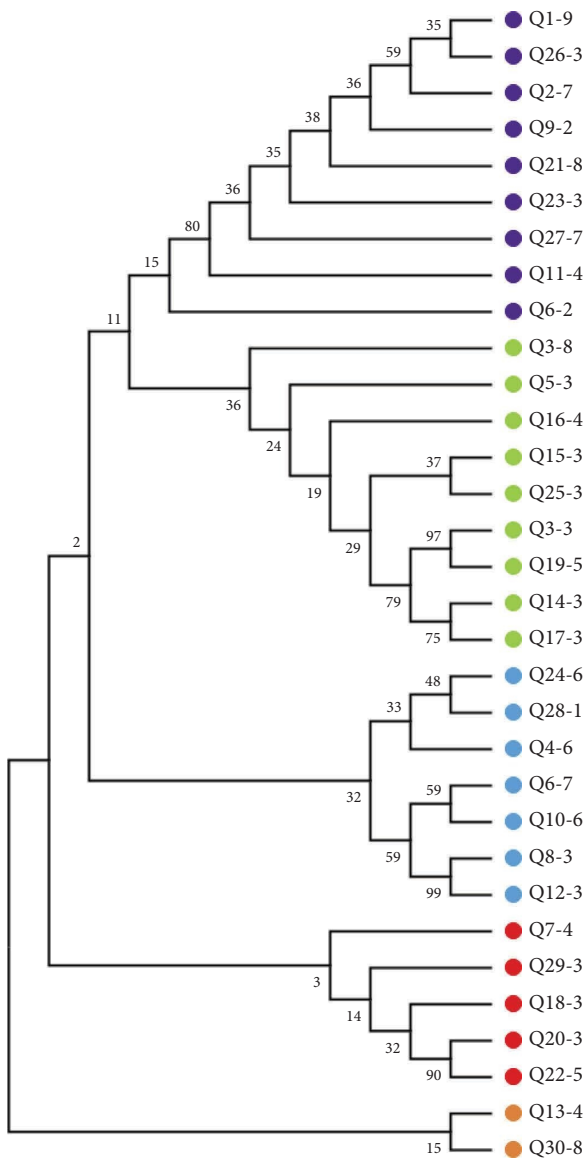

Figure 3: Phylogenetic tree based on the combination of *matK* and *rbcL* sequences of 32 quinoa accessions by the neighbor-joining method.

similarities ranged from 80% to 81%. The evolutionary divergence in *Chenopodium* using *rbcL* gene sequences for 19 accessions found 0.68% diversity in interspecific sequences [6].

Studies on the *Chenopodium* genus have shown the ability of a single barcoding gene (*rbcL*) or a combination of these genes to distinguish individuals of the *Chenopodium* genus [15, 35]. In addition, phylogeny factors and the identification of associated gene families are essential tools for understanding the evolution, mutation, and genetic factors associated with the main characteristics of the *Chenopodium* genus [25].

For the conservation or genetic improvement of plants, phylogenetic studies represent the basis for inferring the evolutionary history of the species, their delimitation, genetic differentiation, and gene flow [36]. This information provides the necessary tools for conservation and input to predict the genetic diversity of species in breeding programs [37]. The reconstruction of the phylogenetic trees that took into account the *matK* region, *rbcL* region, and both concatenated regions in this study suggested that most of the quinoa accessions are closely related, which implies a close relationship that is associated with the reproductive system of the species and its coevolution processes in the Andean productive systems. Therefore, our results support future quinoa breeding and genetic conservation programs.

5. Conclusions

The present study demonstrated the efficacy of the *matK* and *rbcL* genes as reliable markers to support *Chenopodium quinoa* conservation and improvement programs, locally, nationally, and internationally. Although the results of the *rbcL* gene were less promising, the use of both genetic markers for the molecular identification of the *Chenopodium* genus is recommended as an effective, modern, and widely used method.

The phylogenetic trees obtained from the *matK* and *rbcL* gene analyses provided strong evidence of phylogenetic relationships between the evaluated accessions, where similar clades were established between the phylogenetic trees obtained with the *matK* gene and the tree where both *matK* and *rbcL* genes were concatenated. This finding implies that genetic divergence in quinoa is phylogenetically informative at the species level. In addition, the low rates of genetic differentiation may be due to the lack of certified seeds and the constant exchange of seeds between farmers in the principal producing areas of the Andean region. Therefore, DNA barcoding in *Chenopodium quinoa* is an appropriate tool for establishing the correct identification of this species, although more phylogenetic studies involving more species from the *Chenopodium* genus are needed to fully understand the phylogeny of quinoa. The composition and sequence data provide a database for other studies on quinoa diversity.

Data Availability

The data used for the elaboration of this manuscript are available upon request. All sequence data were uploaded to the National Center for Biotechnology Information (NCBI); the access codes of the 62 sequences are OQ148485-OQ148548 (https://www.ncbi.nlm.nih.gov/popset/?term=OQ148485).

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the "Patrimonio Autónomo, Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas-Min-Ciencias. Cód. 63924," to the CIDE group "Competitividad, Innovación y Desarrollo Empresarial," the Faculty of Agricultural Sciences of the Universidad Pedagógica y Tecnológica de Colombia.

Supplementary Materials

Table S1: nucleotide composition of *matK* and *rbcL* chloroplast genes of quinoa accessions from Colombia. (*Supplementary Materials*)

References

[1] E. H. El-Harty, A. Ghazy, T. K. Alateeq et al., "Morphological and molecular characterization of quinoa genotypes," *Agriculture*, vol. 11, no. 4, p. 286, 2021.

[2] M. H. Sellami, C. Pulvento, and A. Lavini, "Agronomic practices and performances of quinoa under field conditions: a systematic review," *Plants*, vol. 10, no. 1, pp. 72–20, 2020.

[3] E. H. Manjarres-Hernández, D. M. Arias-Moreno, A. C. Morillo-Coronado, Z. Z. Ojeda-Pérez, and A. Cárdenas-Chaparro, "Phenotypic characterization of quinoa (Chenopodium quinoa Willd.) for the selection of promising materials for breeding programs," *Plants*, vol. 10, no. 7, pp. 1339–1416, 2021.

[4] M. Romero, A. Mujica, E. Pineda, Y. Ccamapaza, and N. Zavalla, "Genetic identity based on simple sequence repeat (SSR) markers for quinoa (Chenopodium quinoa Willd.)," *Agricultural Science and Research*, vol. 46, no. 2, pp. 166–178, 2019.

[5] J. Salazar, M. De Lourdes Torres, B. Gutiérrez, and A. F. Torres, "Molecular characterization of Ecuadorian quinoa (Chenopodium quinoa Willd.) diversity: implications for conservation and breeding," *Euphytica*, vol. 215, no. 3, p. 60, 2019.

[6] R. J. Devi and N. K. Chrungoo, "Evolutionary divergence in Chenopodium and validation of SNPs in chloroplast rbcL and matK genes by allele-specific PCR for development of Chenopodium quinoa-specific markers," *The Crop Journal*, vol. 5, no. 1, pp. 32–42, 2017.

[7] S. Afiah, W. Hassan, and A. Al Kady, "Assessment of six quinoa (Chenopodium quinoa Willd.) genotypes for seed yield and its attributes under toshka conditions," *Zagazig Journal of Agricultural Research*, vol. 45, no. 6, pp. 2281–2294, 2018.

[8] M. X. Wang, X. T. Wu, J. W. Zou et al., "Genome-wide microsatellite characterisation and marker development in *Chenopodium quinoa*," *Annals of Applied Biology*, vol. 175, no. 3, pp. 415–423, 2019.

[9] M. F. Arif, G. R. Aristya, D. Subositi, A. N. Sari, and R. S. Kasiamdari, "Short Communication: rbcL and matK chloroplast DNA composition of green chireta (Andrographis paniculata) from Indonesia," *Biodiversitas Journal of Biological Diversity*, vol. 20, no. 12, pp. 3575–3583, 2019.

[10] T. Fang, S. Liao, X. Chen et al., "Forensic drowning site inference employing mixed pyrosequencing profile of DNA barcode gene (rbcL)," *International Journal of Legal Medicine*, vol. 133, no. 5, pp. 1351–1360, 2019.

[11] V. T. Ho, T. K. P. Tran, T. T. Vu, and S. Widiarsih, "Comparison of matK and rbcL DNA barcodes for genetic classification of jewel orchid accessions in Vietnam," *Journal of Genetic Engineering and Biotechnology*, vol. 19, no. 1, p. 93, 2021.

[12] D. M. Shadrin, "DNA barcoding: applications," *Russian Journal of Genetics*, vol. 57, no. 4, pp. 489–497, 2021.

[13] A. K. Wang, Q. F. Lu, Z. X. Zhu et al., "Exploring phylogenetic relationships within the subgenera of Bambusa based on DNA barcodes and morphological characteristics," *Scientific Reports*, vol. 12, no. 1, p. 8018, 2022.

[14] P. Maughan, L. Chaney, D. Lightfoot et al., "Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (Chenopodium quinoa Willd.)," *Scientific Reports*, vol. 9, no. 1, p. 185, 2019.

[15] S. Ibrahim, A. El-Hakim, H. E. Ali, and A. El-Maksoud, "Genetic differentiation using ISSR, SCoT and DNA Barcoding for quinoa genotypes," *Arab Journal of Biotechnology*, vol. 22, no. 2, pp. 1–15, 2019.

[16] M. P. Pham, V. H. Tran, D. D. Vu, Q. K. Nguyen, and S. N. Shah, "Phylogenetics of native conifer species in Vietnam based on two chloroplast gene regions rbcL and matK," *Czech Journal of Genetics and Plant Breeding*, vol. 57, no. 2, pp. 58–66, 2021.

[17] M. L. Hollingsworth, A. Andra Clark, L. L. Forrest et al., "Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants," *Molecular Ecology Resources*, vol. 9, no. 2, pp. 439–457, 2009.

[18] H. D. Huynh, T. G. Nguyen, H. X. Duong et al., "Using some DNA barcode for the genetic analysis and identifying some species of Anoectochilus spp," *Can Tho University Journal of Science*, vol. 55, no. 1, pp. 14–23, 2019.

[19] P. Cuenoud, V. Savolainen, L. W. Chatrou, M. Powell, R. J. Grayer, and M. W. Chase, "Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences," *American Journal of Botany*, vol. 89, no. 1, pp. 132–144, 2002.

[20] O. Ihuoma, "Quinoa genetics," in *Biology and Biotechnology of Quinoa*, A. Varma, Ed., Springer, Berlin, Germany, 2021.

[21] A. C. M. Coronado, E. H. M. Hernández, and Y. M. Coronado, "Phenotypic diversity of agromorphological characteristics of quinoa (Chenopodium quinoa Willd.) germplasm in Colombia," *Agricultural Science*, vol. 79, no. 4, 2022.

[22] S. L. Dellaporta, J. Wood, and J. B. Hicks, "A plant DNA minipreparation: version II," *Plant Molecular Biology Reporter*, vol. 1, no. 4, pp. 19–21, 1983.

[23] J. C. Teixeira and C. D. Huber, "The inflated significance of neutral genetic diversity in conservation genetics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 10, 2021.

[24] E. H. Manjarres-Hernández and A. C. Morillo-Coronado, "Genetic diversity of Colombian quinoa (Chenopodium quinoaWilld.): implications for breeding programs," *Genetic Resources and Crop Evolution*, vol. 69, no. 7, pp. 2447–2458, 2022.

[25] K. Li, Y. Fan, G. Zhou et al., "Genome-wide identification, phylogenetic analysis, and expression profiles of trihelix transcription factor family genes in quinoa (Chenopodium quinoa Willd.) under abiotic stress conditions," *BMC Genomics*, vol. 23, no. 1, p. 499, 2022.

[26] R. P. Srivastava, G. Saxena, L. Singh, A. Singh, P. C. Verma, and G. Kaur, "Interspecific and intraspecific analysis of Selinum spp. collected from Indian Himalayas using DNA barcoding," *Journal of Genetic Engineering and Biotechnology*, vol. 20, no. 1, p. 63, 2022.

[27] A. A. Algarni, "Molecular identification and phylogenetic analysis of Aloe shadensis from Saudi Arabia based on matK, rbcL and ITS DNA barcode sequence," *Saudi Journal of Biological Sciences*, vol. 29, no. 2, pp. 1125–1133, 2022.

[28] N. Kipkiror, E. K. Muge, D. M. Ochieno, and E. Nyaboga, "DNA barcoding markers provide insight into species discrimination, genetic diversity and phylogenetic relationships of yam (Dioscorea spp.)," *Biologia*, vol. 78, no. 3, pp. 689–705, 2022.

[29] H. J. De Boer, A. Ghorbani, V. Manzanilla et al., "DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade," *Proceedings of the Royal Society B: Biological Sciences*, vol. 284, no. 1863, Article ID 20171182, 2017.

[30] C. Guo, Y. Luo, L. M. Gao et al., "Phylogenomics and the flowering plant tree of life," *Journal of Integrative Plant Biology*, vol. 65, no. 2, pp. 299–323, 2023.

[31] H. T. Li, Y. Luo, L. Gan et al., "Plastid phylogenomic insights into relationships of all flowering plant families," *BMC Biology*, vol. 19, no. 1, p. 232, 2021.

[32] S. Fuentes-Bazan, G. Mansion, and T. Borsch, "Towards a species level tree of the globally diverse genus Chenopodium (Chenopodiaceae)," *Molecular Phylogenetics and Evolution*, vol. 62, no. 1, pp. 359–374, 2012.

[33] Y. Kang, Z. Deng, R. Zang, and W. Long, "DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests," *Scientific Reports*, vol. 7, no. 1, Article ID 12564, 2017.

[34] X. Huang, X. Ci, J. G. Conran, and J. Li, "Application of DNA barcodes in Asian tropical trees-a case study from Xishuangbanna Nature Reserve, southwest China," *PLoS One*, vol. 10, no. 6, 2015.

[35] M. Gao, Y. Dong, V. Valcárcel, Z. Ren, and Y. Li, "Complete chloroplast genome of the grain Chenopodium quinoa Willd., an important economical and dietary plant," *Mitochondrial DNA Part B*, vol. 6, no. 1, pp. 40–42, 2021.

[36] D. H. Mursyidin, M. I. Makruf, A. Badruzsaufari, and A. Noor, "Molecular diversity of exotic durian (Durio spp.) germplasm: a case study of Kalimantan, Indonesia," *Journal of Genetic Engineering and Biotechnology*, vol. 20, no. 1, p. 39, 2022.

[37] G. Acquaah, *Principles of Plant Genetics and Breeding*, Blackwell Publishing, Oxford, UK, 2007.