BMC Plant Biology

RESEARCH Open Access

Comparative chloroplast genome analyses of cultivated and wild *Capsicum* species shed light on evolution and phylogeny

Raveendar Sebastin[1†], Jaewook Kim[2†], Ick-Hyun Jo[3], Ju-Kyung Yu[4], Woojong Jang[5], Seahee Han[6], Hyun-Seung Park[7], Amal Mohamed AlGarawi[8], Ashraf Atef Hatamleh[8], Yoon-Sup So[4], Donghwan Shim[9*] and Jong-Wook Chung[1*]

Abstract

The chloroplast (cp.) genome, also known as plastome, plays crucial roles in plant survival, adaptation, and evolution. The stable genetic structure of cp. genomes provides an ideal system for investigating species evolution. We sequenced three complete cp. genome sequences of Capsicum species and analyzed them using sequences of various Capsicum species retrieved from the NCBI database. The cp. genome of *Capsicum* species maintains a well-preserved quadripartite structure consisting of two inverted repeats (IRs) flanked by a large single copy (LSC) region and a small single copy (SSC) region. The sizes of cp. genome sequences ranged from 156,583 bp (*C. lycianthoides*) to 157,390 bp (*C.pubescens*). A total of 127–132 unique genes, including 83–87 protein-coding, 36–37 tRNA, and eight rRNA genes, were predicted. Comparison of cp. genomes of 10 *Capsicum* species revealed high sequence similarity in genome-wide organization and gene arrangements. Fragments of *trnT-UGU/trnL-UAA*, *ccsA*, *ndhD*, *rps12*, and *ycf1* were identified as variable regions, and nucleotide variability of LSC and SSC was higher than that of IR. Phylogenetic speciation analysis showed that the major domesticated *C. annuum* species were the most extensively divergent species and closely related to *C. tovarii* and *C. frutescens*. Analysis of divergent times suggested that a substantial range of speciation events started occurring ~ 25.79 million years ago (Mya). Overall, comparative analysis of cp. genomes of *Capsicum* species not only offers new insights into their genetic variation and phylogenetic relationships, but also lays a foundation for evolutionary history, genetic diversity, conservation, and biological breeding of *Capsicum* species.

Keywords *Capsicum*, Chloroplast, Phylogeny, Speciation, Species evolution, Taxonomy

[†]Raveendar Sebastin and Jaewook Kim contributed equally to this work.

*Correspondence:
Donghwan Shim
dshim104@cnu.ac.kr
Jong-Wook Chung
jwchung73@chungbuk.ac.kr
[1]Department of Industrial Plant Science and Technology, Chungbuk National University, Cheongju 28644, Republic of Korea
[2]Department of Biology Education, Korea National University of Education, Cheongju 28173, Republic of Korea
[3]Department of Crop Science and Biotechnology, Dankook University, Cheonan 31116, Republic of Korea
[4]Department of Crop Science, Chungbuk National University, Cheongju 28644, Republic of Korea
[5]Herbal Medicine Resources Research Center, Korea Institute of Oriental Medicine, Naju 58245, Republic of Korea
[6]Honam National Institute of Biological Resources, Mokpo 58762, Republic of Korea
[7]Department of Integrative Biological Sciences and Industry, Convergence Research Center for Natural Products, Sejong University, Seoul 05006, Republic of Korea
[8]Department of Botany and Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
[9]Department of Biological Science, Chungnam National University, Daejeon 34134, Republic of Korea

Introduction

Pepper (*Capsicum* spp.) is the most economically significant genus within the family Solanaceae and contains approximately 35 species, of which five major species represent ancient domestication [1]. The production and usage of *Capsicum* has spread worldwide across civilizations, which subsequently led to selective pressure associated with cultivation under different environmental conditions [2]. According to the recent data (FAOSTAT; http://faostat.fao.org/), worldwide pepper production surpassed 3.68 million hectares (Mhs), yielding a total production of 36.29 million tons (Mts) of green peppers and 4.84 Mts of dried peppers. Pepper fruits are used in food and non-food processing applications in many innovative ways [3–5]. The genus *Capsicum* comprises domesticated and wild species with highly variable morphologies. Systemic classification of the *Capsicum* genus was achieved by considering morphology, cytogenetics, biochemistry, and ethnobotanical data [6]. Recently, approximately 34 of the 35 species were recognized using comprehensive phylogenetic analysis [7]. In subsequent iterations of the analysis, 11 major clades were meticulously resolved within the *Capsicum* genus. These clades were then assigned informal designations to aid their identification and characterization [8, 9].

Indeed, archaeological evidence indicates separate domestication events among cultivated pepper species occurred as early as 6000 years ago [10]. Pepper made its voyage from the West Indies to Europe during the late 15th and early 16th centuries, and its rapid distribution to Africa and Asia, including China, holds the earliest documentation of pepper [11, 12]. The course of domestication and controlled breeding yielded non-deciduous pepper cultivars exhibiting a wide array of morphologically and agronomically important traits, including diverse fruit shapes, varying sizes, distinct weights, pendulous fruit orientation, as well as a broad spectrum of pungency levels [13]. More recently, large, sweet, and non-pungent bell peppers have emerged through selective breeding, with the earliest records dating to the 1700s [14].

Capsicum annuum, commonly known as pepper, has been extensively bred and is primarily considered in breeding programs to develop genetically superior genotypes. Selective pressure leads to the advent of genomic selection signals, which are quantified as decreases in nucleotide diversity [15]. Several genomic selection signals have been detected in the pepper genome owing to early domestication [16]. Initial evolutionary shifts within the slender-fruited pepper category were marked by selective pressure acting on extensive assemblages of candidate genes implicated in determination of fruit pungency and/or shape. This selective phenomenon was likely facilitated by the considerable genetic diversity existing within these groups, which was amplified by the absence of genetic bottlenecks [16, 17]. Conversely, the subsequent genomic transition exhibited targeted selection towards a narrower subset of genes. This observation supports the hypothesis that genetic bottlenecks may occur during this transition. This bottleneck scenario is plausibly attributable to the transport of a subset of gene pools via the historical Silk Road route towards East Asia [16].

Wild *Capsicum* species such as *C. chinense*, *C. frutescens*, *C. baccatum*, and *C. pubescence* are well-known for their intercrossing characteristics with *C. annuum* to produce fertile progenies. Thus, to enhance breeding efficiency of *C. annuum*, inter-breeding of *C. annuum* with other wild *Capsicum* species has been performed [18]. Several agronomic and physiological traits found in wild *Capsicum* species are considered valuable genetic sources to improve the characteristics of C. *annuum* [19]. Especially, capsaicinoids comprise a group of substances produced only within the *Capsicum* genus, which is one of the pivotal criteria for hot pepper breeding [20]. Research on various *Capsicum* species is not only relevant to improve *C. annuum*, but also to discover novel species. This may lead to domestication and preservation of their genetic pool in *Capsicum* species and may play a significant role in conservation of *Capsicum* species.

Despite the economic importance of *C. annuum*, the availability of wild *Capsicum* species for use in cultivar breeding is relatively limited within the Solanum family. Recent advances in next-generation sequencing (NGS) technology have made whole-genome sequencing and analysis much more straightforward than in the past. Whole genome-scale datasets encompass a larger number of single nucleotide polymorphisms (SNPs), which can significantly aid in reconstructing well-supported species-specific phylogenetic trees [21]. However, the large genome size of *Capsicum* species (>3 Gbp) with long terminal repeat (LTR) sequences hinders its genomic analysis [22–24]. Plastid genome assembly and analysis are relatively simple approaches for analyzing the phylogenetic history of a species [25] and are highly useful in resolving taxonomic complexes within and between species and provides insights for breeding programs [26]. In particular, maternal inheritance of plastids provides stable molecular evidence for each species.

Despite the presence of well-annotated genomic sequences for numerous pepper species, our comprehension of the molecular evolutionary trajectory of this crop trails behind that of its closely related counterpart, Solanum. To fill this gap, we sequenced three wild *Capsicum* species, and compared their genomes to other publicly available plastid genomes of *Capsicum* species. The findings of this study may provide new guidelines for evolutionary framework of *Capsicum* species and facilitate the development and utilization of its genetic resources.

Materials and methods

Plant materials and DNA extraction

Seeds of *C. chacoense*, *C. galapagoense*, and *C. tovarii* (Accession Nos. 261663, 236750, and K057567, respectively) were obtained from the National Agrobiodiversity Center, Republic of Korea. These seeds were germinated in a greenhouse and fresh leaves from 40-day-old seedlings were ground in liquid nitrogen using a mortar and pestle. Genomic DNA was extracted using the Exgene Plant SV midi kit (GeneAll Biotechnology, Seoul, Korea) following the manufacturer's instructions. The quantity and quality of the extracted DNA were measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific Inc., Wilmington, DE, USA).

Genome sequencing, assembly, and annotation

An Illumina paired-end DNA library with an average insert size of 500 base pairs (bp) was constructed using the Illumina TruSeq library preparation kit, following the manufacturer's instructions. Subsequently, the library was sequenced using the MiSeq instrument at LabGenomics (http://www.labgenomics.co.kr/) to generate paired-end reads of 300 bp each. Complete cp. genome sequences were assembled through *de novo* assembly of low-coverage whole genome sequence (WGS) data. Low-quality sequences (quality score < 20, Q20) were initially filtered out, and the remaining high-quality reads were assembled using the CLC Genome Assembler (version beta 4.6, CLC Inc., Aarhus, Denmark), employing an overlap size ranging from 200 to 600 bp. Next, chloroplast contigs were identified within the initial assembly using BLAST (ver. 2.2.31) search against reference cp. (NCBI accession NC_018552) sequences. The selected contigs were oriented to construct the final complete cp. genome structure, and manual corrections were made to address any ambiguous nucleotides or gaps.

The assembled chloroplast genome was annotated using the web-based program Dual Organellar GenoMe Annotator (DOGMA; http://dogma.ccbb.utexas.edu/) with default parameters. This annotation encompasses the prediction of protein-coding, transfer RNA (tRNA), and ribosomal RNA (rRNA) genes. Next, the positions of intron-containing genes were identified using BLASTN (ver. 2.2.31) alignment of the assembled cp. genome against a reference cp. genome. A graphical map illustrating the arrangement of genes within the cp. genome was generated using CPGView (http://www.1kmpg.cn/cpgview/; Accessed in July 2023). Additionally, cp. genomes of 19 *Capsicum* species were retrieved from the NCBI database for subsequent comparative analysis (Supplementary Table S1).

Codon usage analysis

A total of 87 inherited protein-coding genes within three *Capsicum* cp. genomes of *C. chacoense*, *C. galapagoense* and *C. tovarii* were concatenated to facilitate computation of synonymous codon usage analysis. Number of codons (Nc) and relative synonymous codon usage (RSCU) were calculated using the CodonW1.4.2 (Available online: http://downloads.fyxm.net/CodonW-76666.html) program. The Nc, present within individual genes, serves as a straightforward gauge of codon bias [27], with Nc values spanning a spectrum ranging from 20 (indicative of extreme bias) to 61 (totally unbiased). RSCU is the ratio of observed frequency of a specific codon to its anticipated frequency [28]. In situations where synonymous codons, each encoding the same amino acid, were utilized with equal frequency, the RSCU values were approximately 1.0, signifying an absence of bias. In addition, the frequency distribution of amino acids (AA) was quantified and is represented as the percentage of codons encoding a particular amino acid relative to the total codon count.

Comparative analysis of cp. Genomes

Expansion/contraction events in the cp. genomes of 10 *Capsicum* species (Table 1) were compared using IRScope (https://irscope.shinyapps.io/irapp/). Complete cp. genome sequences of *Capsicum* species were compared using the mVISTA online tool (http://genome.lbl.gov/vista/mvista/submit.shtml). This analysis was performed using the Shuffle-LAGAN mode [29]. To determine nucleotide diversity in the cp. genomes, the 10 *Capsicum* cp. genome sequences were aligned using the program MAFFT 7.388 [30]. Ambiguously aligned loci ('N') were excluded from the analyses. Unambiguously aligned DNA sequences were used for the nucleotide diversity analysis. To identify hypervariable sites among *Capsicum* chloroplast genomes, a sliding window analysis was performed using DnaSP v6 software [31]. This analysis utilized a window length of 600 base pairs and a step size of 200 base pairs to calculate the nucleotide diversity (π).

Phylogenetic analysis and divergence time estimation

Twenty-two cp. genomes of *Capsicum* species, including three newly sequenced cp. genomes of species such as *C. chacoense*, *C. galapagoense*, and *C. tovarii*, were used to construct Maximum Likelihood (ML) phylogenic trees employing the Kimura 2-parameter model (with 1000 bootstrap repeats) using MEGA X [32] after aligning whole cp. genome sequences using MAFFT 7.388. The divergence times among *Capsicum* species were

Table 1 Summary statistics for the sequencing and assembly of nine Capsicum species chloroplast genomes

Genome features	C. chacoense	C. galapagoense	C. tovarii	C. annuum	C. baccatum	C. chinense	C. eximium	C. frutescens	C. pubescens	C. lycianthoides
NCBI No.	NC_033525	NC_033524	KX913219	KR078313	KR078314	KX913217	KX913220	KR078312	NC_039694	NC_026551
Total cpDNA size (bp)	156,995	156,959	156,816	156,878	157,145	156,936	156,947	156,817	157,390	156,583
LSC region (bp)	87,379	87,347	87,379	87,347	87,351	87,330	87,341	87,380	87,688	86,813
IRs region (bp)	25,859	25,847	25,792	25,801	25,910	25,847	25,847	25,792	25,887	25,624
SSC region (bp)	17,898	17,918	17,853	17,929	17,974	17,912	17,912	17,853	17,928	18,522
Total GC content (%)	37.68	37.71	37.72	37.72	37.72	37.72	37.72	37.72	37.69	37.76
Total number of genes	132	132	132	131	131	132	132	131	131	127
Protein encoding genes	87	87	87	86	86	87	87	86	86	83
tRNA genes	37	37	37	37	37	37	37	37	37	36
rRNA genes	8	8	8	8	8	8	8	8	8	8
No. raw reads	9,641,714	7,563,398	8,644,550	-	-	-	-	-	-	-
Total read length (bp)	2,893,962,655	2,269,579,470	2,592,924,809	-	-	-	-	-	-	-
No. mapped reads	151,526	138,398	115,885	-	-	-	-	-	-	-
Cp genome coverage (x)	206	184	156	-	-	-	-	-	-	-

estimated using the Bayesian method implemented in the BEAST program [33]. In BEAST, the lengths of the branches are converted into units of evolutionary time. Essentially, this allows the tree to be placed in a time-based context for a better understanding of evolutionary processes.

In the present study, an uncorrelated relaxed lognormal clock model was employed to enable independent optimization of evolutionary rates for each branch in the phylogenetic tree. The speciation process was represented using the Yule process [34]. Branching rates were determined using the GTR+I+G nucleotide substitution model, and two separate Markov chain Monte Carlo (MCMC) runs were conducted using BEAST, each spanning 100,000,000 generations. Sampling was performed at intervals of 10,000 generations during these runs to ensure comprehensive exploration of the parameter space and convergence of the analysis.

To calibrate the phylogenetic tree of cp. genomes for the 22 *Capsicum* species accessions, we incorporated estimated dates of speciation events between *Capsicum* and Solanum species, which were sourced from www.timetree.org. These events were indicated to have occurred approximately 21.3 million years ago (Mya). In our study, *Solanum lycopersicum* and *Lycianthes radiata* were designated as outgroup species, and age of the tree roots was constrained by a normal distribution, with a mean of 21.38 Mya. To further analyze the results generated by BEAST, we used Tracer v. 1.6 [35] to explore the output, enabling us to assess various statistical aspects of the phylogenetic analysis. Additionally, TreeAnnotator v.1.8, a component of the BEAST package, was employed to generate maximum clade credibility trees from the post-burn-in trees. This process facilitated determination of 95% posterior density of ages for all nodes within the phylogenetic tree, thus providing crucial insights into the timeline of speciation events among *Capsicum* species. To visualize the outcomes, maximum clade credibility trees were annotated using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/, Accessed on July 18, 2023). Principal component analysis (PCA) was conducted on all samples utilizing TASSEL 5.0 [36], and the resulting data were visualized using XLSTAT software (www.xlstat.com, Accessed on July 10, 2024).

Results

Plastid genome size and Organization

This study focused on analysis of cp. genomes of three *Capsicum* species that are wild relatives of domesticated *Capsicum* species. The HiSeq2500 generated approximately 7.75 GB of paired-end sequence data. The plastid genome coverage for *C. chacoense*, *C. galapagoense*, and *C. tovarii* with the reference genome reached 206x, 184x, and 156x, respectively. The cp. genomes of the three *Capsicum* species, along with their sizes, reads, GC contents, and averages are shown in Table 1. We obtained complete cp. genome maps of *C. chacoense*, *C. galapagoense*, and *C. tovarii* through a *de novo* assembly followed by reference-guided assembly using the *C. annuum* (NC_018552) cp. genome as a reference (Fig. 1A, B, and C). Size of the *Capsicum* cp. genome was 156,995 bp in *C. chacoense*, 156,959 bp in *C. galapagoense*, and 156,818 bp in *C. tovarii* (Table 1). The complete cp. genome size of *Capsicum* species was very similar to that of the reference genome (approximately 156 kb).

Combined with the newly assembled cp. genomes, we collected 19 available cp. genomes for *Capsicum* species, of which 10 were chosen to characterize genomic features of inter-species cp. genome (Table 1). The lengths of

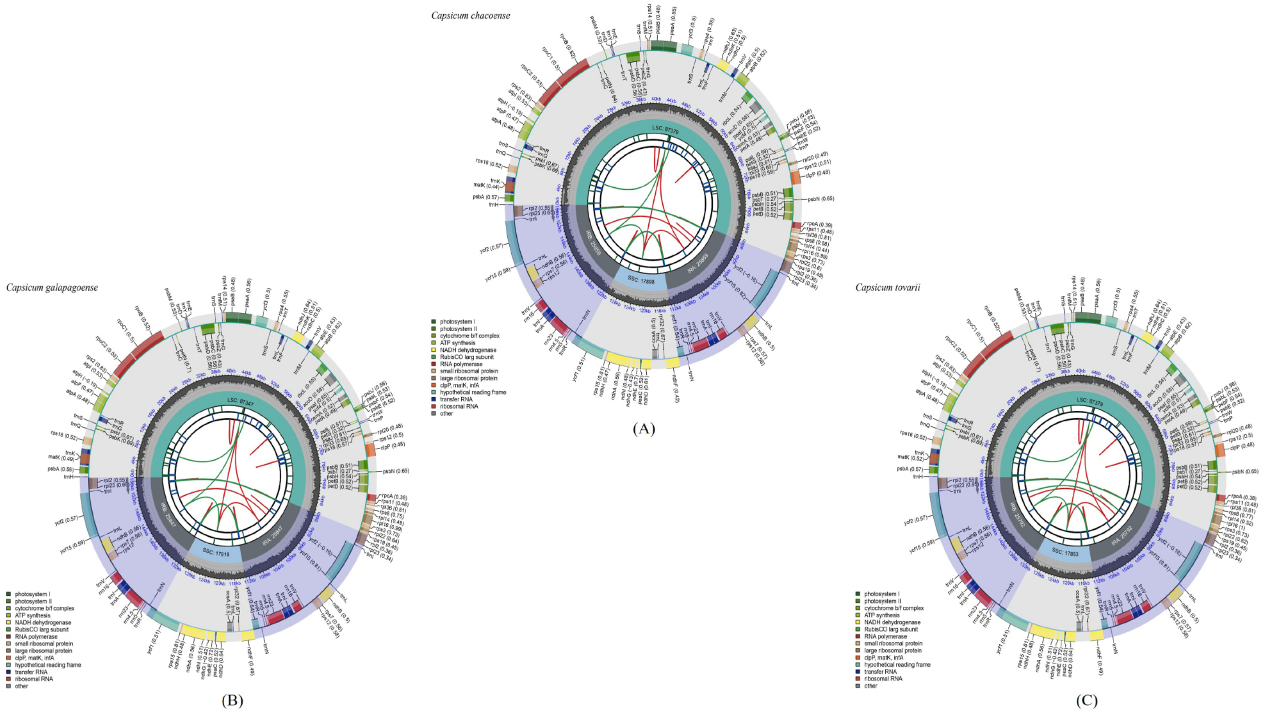

Fig. 1 Structure of the cp. genomes of three *Capsicum* species (*C. chacoense*, *C. galapagoense*, and *C. tovarii*). The map consists of six tracks. Starting from the center, the first track shows scattered repeats (direct and palindromic) linked by red and green arcs. The second track displays long tandem repeats as blue bars, while the third track shows short tandem repeats (microsatellites) as differently colored short bars based on their sizes. Colors correspond to different repeat sizes: black for complex repeats, green (size 1), yellow (size 2), purple (size 3), blue (size 4), orange (size 5), and red (size 6). The fourth track marks specific regions of the genome. The fifth track depicts the GC content, and the sixth track highlights the genes, named by codon usage bias and colored by their functions. Inner genes are transcribed clockwise, and outer genes are transcribed anticlockwise. The bottom left corner denotes the gene functions

the 10 cp. genomes varied from 156,583 (*C.lycianthoides*) to 157,390 bp (*C.pubescens*). As reported previously, all the 10 *Capsicum* cp. genomes contained two copies of IR regions, each ranging from 25,792 (*C. tovarii* and *C. frutescens*) to 25,887 bp (*C.pubescens*). The IRs were separated by an LSC region ranging from 87,330 (*C. chinense*) to 87,688 bp (*C.pubescens*) and an SSC region ranging from 17,853 (*C. tovarii* and *C. frutescens*) to 17,929 bp (*C. annuum var. annuum*). The overall GC content in all the 10 *Capsicum* cp. genomes was similar to that reported previously, with an average of 37.71% (Table 1).

The three *Capsicum* cp. genomes sequenced in this study contained 132 genes, comprising 87 protein-coding genes (CDS), eight rRNAs, and 37 tRNAs (Table 2). Seventeen genes containing introns were identified in the cp. genome. Among the annotated genes, 14 *(trnA-UGC*, *trnG-UCC*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *rps16*, *rpl2*, *rpl16*, *rpoC1*, *ndhA*, *ndhB*, *petB*, *petD*, and *atpF*) contained one intron each, and three (*rps12*, *ycf3*, and *clpP*) contained two introns each (Table 2). In addition, 19 genes were determined to be duplicated, of which three genes such as *rps19*, *rpl2*, and *rpl23* were found to have one or two copies. The complete cp. genome sequences with gene annotations have been submitted to NCBI under GenBank accession numbers NC_033525 for *C. chacoense*, NC_033524 for *C. galapagoense*, and KX913219 for *C. tovarii.*

Codon usage analysis

Codon usage frequency of cp. genomes of three *Capsicum* species such as *C. chacoense*, *C. galapagoense* and *C. tovarii* was calculated based on protein-coding genes (CDSs). The number of codons within the protein-coding genes of the three plastid genomes ranged from 26,945 (*C. chacoense*) to 26,991 (*C. tovarii*). In all three investigated species, leucine (leu; 2908–2957 occurrences) was the most abundant amino acid that was encoded by six codons, followed by isoleucine (ile; 2142–2249 occurrences) that was encoded by three codons (Supplementary Table S2). Conversely, methionine and tryptophan were encoded by a single codon, with occurrences ranging from 593 to 628 and 485–522, respectively. Notably, these amino acids exhibited no codon usage bias, as reflected by their RSCU values of 1. The UUA codon for leucine displayed the highest RSCU values (1.68–1.86), while the AGC codon for serine exhibited the lowest RSCU values (0.37–0.40) across the three cp. genomes. In the cumulative cp. genomes, a set of 30 codons featured RSCU values exceeding one, indicating their preferred usage. Notably, of these 30 codons, 29 culminated in adenine (A) or uracil (U), highlighting a discernible codon preference within the cp. genome (Supplementary Table S2).

Comparative analysis of junction regions

Capsicum plastid genomes displayed a typical quadripartite structure consisting of a pair of IR regions separated by LSC and SSC regions (Fig. 2). Among the 10

Table 2 List of annotated genes in the cp. Genomes of *C. Chacoense*, *C. Galapagoense*, and *C. Tovarii*

Category for genes	Group of genes	Name of genes
Self-replication	rRNA genes	rrn16[a], rrn23[a], rrn4.5[a], rrn5[a]
	tRNA genes	trnA-UGC[*,a], trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC[*], trnH-GUG, trnI-CAU[a], trnI-GAU[a], trnK-UUU[*], trnL-CAA[a], trnL-UAA[*], trnL-UAG, trnM-CAU, trnN-GUU[a], trnP-UGG, trnQ-UUG[a], trnR-ACG[a], trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC[a], trnV-UAC[*], trnW-CCA, trnY-GUA
	Small subunit of ribosome	rps2, rps3, rps4, rps7[a], rps8, rps11, rps12[**,a], rps14, rps15, rps16[*], rps18, rps19[b]
	Large subunit of ribosome	rpl2[*,b], rpl14,rpl16[*], rpl20,rpl22,rpl23[b], rpl32,rpl33,rpl36
	DNA dependent RNA polymerase	rpoA, rpoB, rpoC1[*], rpoC2
Genes for photosynthesis	Subunits of NADH-dehydrogenase	ndhA[*], ndhB[*,a], ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
	Subunits of photosystem I	psaA, psaB, psaC, psaI, psaJ, ycf3[**]
	Subunits of photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	Subunits of cytochrome b/f complex	petA, petB[*], petD[*], petG, petL, petN
	Subunits of ATP synthase	atpA, atpB, atpE, atpF[*], atpH, atpI
	Large subunit of rubisco	rbcL
Other genes	Maturase	matK
	Protease	clpP[**]
	Envelope membrane protein	cemA
	Subunit of Acetyl-CoA-carboxylase	accD
	c-type cytochrome synthesis gene	ccsA
Genes of unknown function	Open Reading Frames (ORF, ycf)	ycf1, ycf2[a], ycf4

[*] Gene with one intron, [**] Gene with two introns, [a] Gene with two copies, [b] Gene with one or two copies

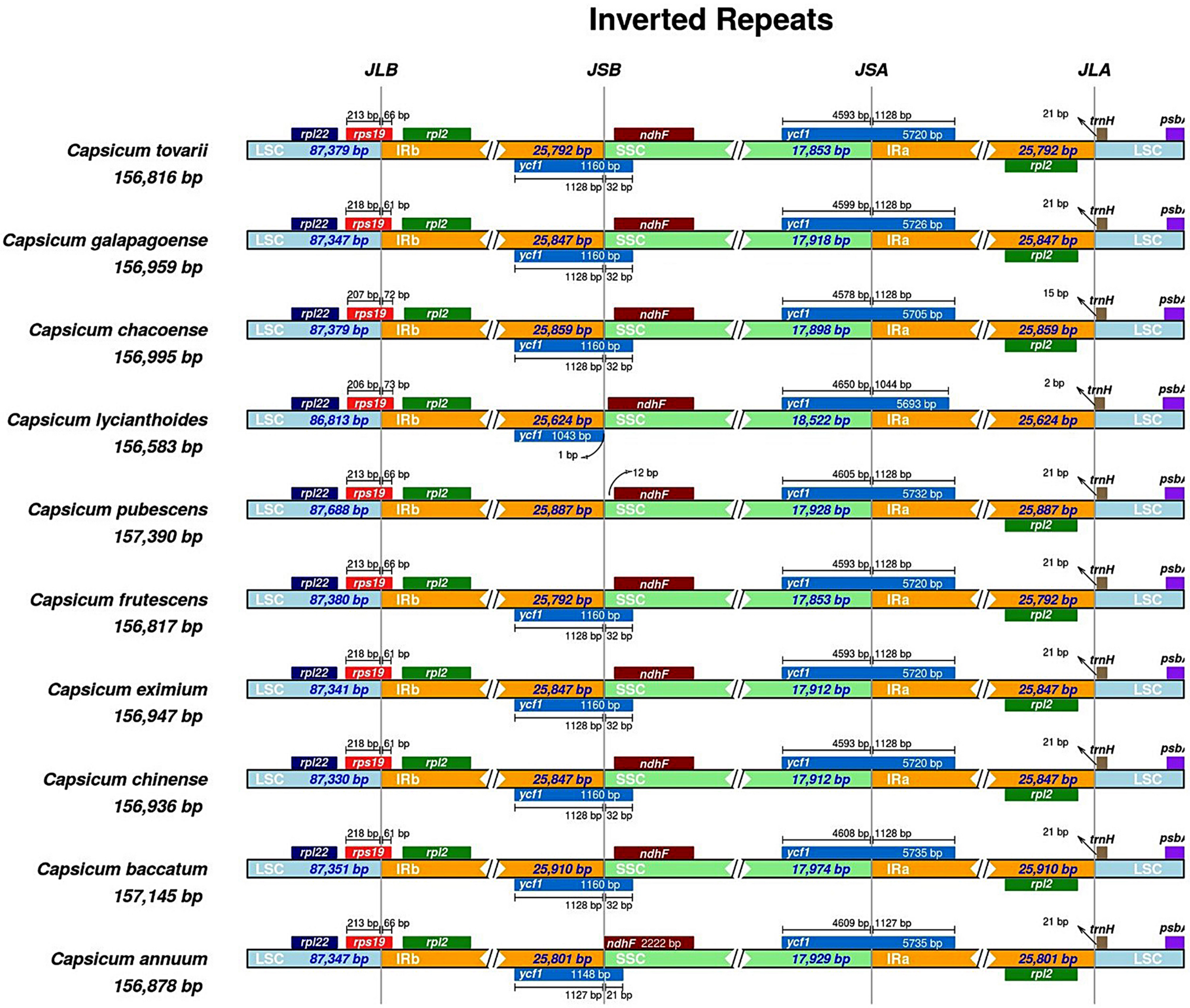

Fig. 2 Comparison of boundaries of LSC, SSC, and IR regions among 10 cp. genomes of *Capsicum* species. The boundary points are labeled as JLB, JSB, JSA, and JLA; each label represents specific junctions related to LSC and IRb, IRb and SSC, SSC and IRa, and IRa and LSC, respectively

species, the SSC region of *C. lycianthoides* was the largest, whereas those of *C. tovarii* and *C. frutescens* were the smallest. Similarly, the IR regions of *C. baccatum* were the largest and those of *C. lycianthoides* were the smallest (Fig. 2). Within all the 10 examined plastid genomes, four distinct boundaries were identified between the regions LSC–IR and SSC–IR. Genes situated in proximity to these boundaries exhibited consistent patterns across all the analyzed chloroplast genomes (Fig. 2). Specifically, regarding the junction between the LSC and IRB regions (JLB), the gene *rps19* was observed to traverse from the LSC region to the IRB region; however, different lengths were observed from 61 (*C. galapagoense*, *C. chinense* and *C. baccatum*) to 73 bp (*C. lycianthoides*) in all the species analyzed (Fig. 2). The *rpl22* and *rpl2* genes, which were adjacent to *rps19* gene at the JLB boundary, were located in the LSC and IRB regions, respectively.

Similarly, for the junction between the SSC and IRB regions (JSB), two genes, *ycf1* and *ndhF*, were placed next to the boundary. The *ndhF* gene was located in the SSC region, whereas *ycf1* was observed to cross over from the SSC region to IRB region, except for *C. pubescens* and *C. lycianthoides*. The *ycf1* gene of *C. pubescens* was absent in the IRB region, and the *ndhF* gene in SSC was also 12 bp away from the boundary. The *ycf1* gene of *C. lycianthoides* was smaller in size (1043 bp) than the other species and was also located 1 bp away from the boundary. By contrast, the *ndhF* gene (2222 bp) in *C. annuum* plastome showed a difference in position and was located adjacent to the JSB boundary; the *ycf1* gene was located approximately 1127 bp in the IRB and 21 bp in SSC region.

Junction analysis between SSC and IRA regions (JSA) demonstrated that *ycf1* was found to be intact at the JSA boundary and crossed over from the SSC into IRA region. However, the *ycf1* gene length in the SSC region varied from 4578 (*C. chacoense*) to 4650 bp (*C. lycianthoides*), whereas in the IRA region, all the species were of 1128 bp, except *C. lycianthoides* (1044 bp) and *C. annuum* (1127 bp). In contrast, junction analysis between IRA and LSC regions (JLA) showed that both the *trnH* and *psbA* genes were placed in the LSC region and a copy of *rpl2* gene was precisely located in the IRA region, except *C. lycianthoides* in which the copy of *rpl2* gene was absent in the IRA region.

Comparative sequence variance analysis

We assessed variations in cp. genomes within congeneric *Capsicum* species using the online alignment tool mVISTA, with *C. frutescens* as a reference (Fig. 3). The results showed that there was no gene rearrangement in the *Capsicum* cp. genome, and the sequence variation in the non-coding regions was higher than that of genes in coding region. LSC and SSC had more diverse loci than IR; several regions were variable and were able to distinguish different populations. Nucleotide polymorphisms (Pi) in the 10 *Capsicum* cp. genomes, were analyzed by sliding window analysis using multiple-sequence alignment, and *C. annuum* showed high sequence similarity (Fig. 4). A total of 1486 common polymorphic sites were detected in the cp. genomes, with a mean nucleotide diversity (Pi) value of 0.00225. The mean nucleotide diversity (Pi) values of the LSC, SSC, and IR regions were 0.00264, 0.00350, and 0.00122, respectively. Four nucleotide polymorphism hotspots (hotspots) were discovered in intergenic regions (*trnT-UGU/trnL-UAA*) and four genes (*ccsA/ndhD/rps12/ycf1*) with Pi values above 0.01 were located in single-copy regions.

Phylogenetic analysis and estimation of divergence time

To investigate phylogenetic relationships of the *Capsicum* species, alignments of all 22 cp. genome sequences were performed using MAFFT, and a phylogenic tree was

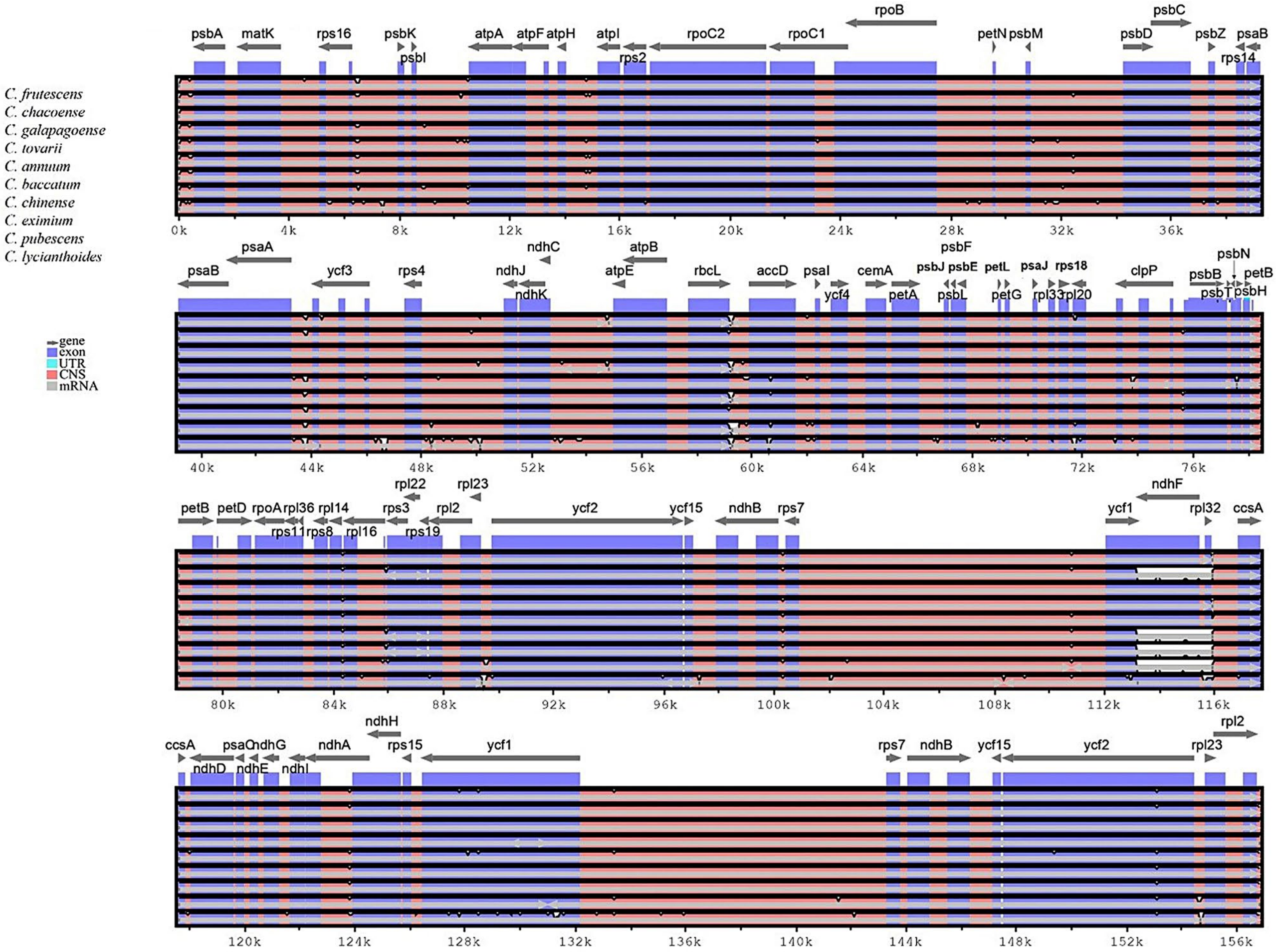

Fig. 3 A sequence similarity plot was generated using the mVISTA tool to compare the cp. genomes of *Capsicum* species. KR078312 (*C. frutescens*) was used as a reference genome. The Y-axis shows percentage of sequence similarity ranging from 50–100%. The direction in which genes are transcribed is indicated by gray arrows. Non-coding sequences (CNS) are depicted as pink bars, while exons are shown as purple bars. Variations in the genome are shown as white peaks

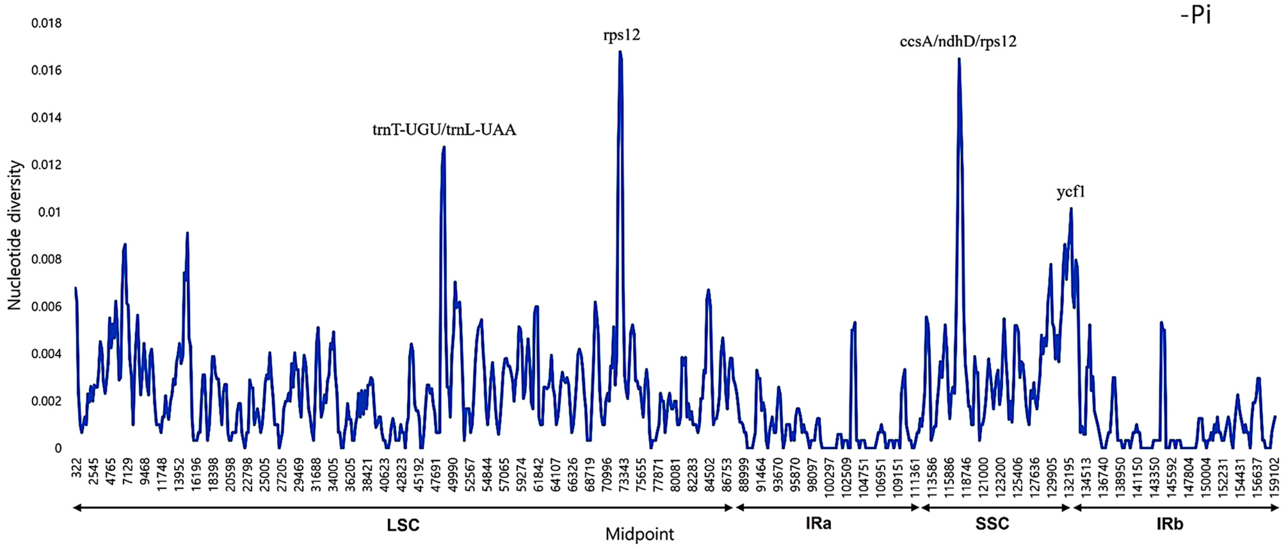

Fig. 4 Sliding window analysis was performed on complete cp. genomes of 10 *Capsicum* species. Nucleotide variability across these genomes was evaluated using DnaSP, employing a window length of 600 base pairs and a step size of 200 base pairs. The X-axis denotes different regions of the cp. genome, whereas the Y-axis represents nucleotide diversity (pi) for each window. Five particularly hypervariable regions, characterized by pi values exceeding 0.01, have been identified and labeled

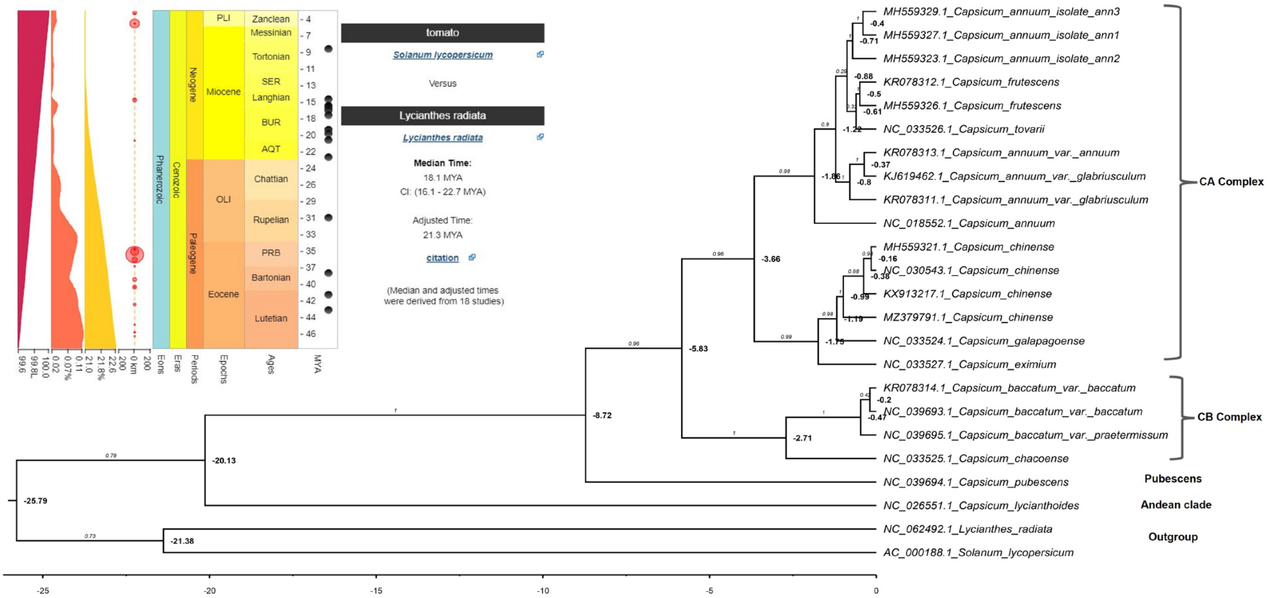
PLI
Zanclean
Messinian
Tortonian
SER
Langhian
BUR
AQT
Chattian
Rupelian
PRB
Bartonian
Lutetian
Miocene
OLI
Eocene
Neogene
Paleogene
Phanerozoic
Cenozoic
- 4
- 7
- 9
- 11
- 13
- 15
- 18
- 20
- 22
- 24
- 26
- 29
- 31
- 33
- 35
- 37
- 40
- 42
- 44
- 46
99.6
99.8L
100.0
0.02
0.07%
0.11
21.0
21.8%
22.6
200
0 km
200
Eons
Eras
Periods
Epochs
Ages
MYA
tomato
Solanum lycopersicum
Versus
Lycianthes radiata
Lycianthes radiata
Median Time:
18.1 MYA
CI: (16.1 - 22.7 MYA)
Adjusted Time:
21.3 MYA
citation
(Median and adjusted times were derived from 18 studies)
MH559329.1_Capsicum_annuum_isolate_ann3
MH559327.1_Capsicum_annuum_isolate_ann1
MH559323.1_Capsicum_annuum_isolate_ann2
KR078312.1_Capsicum_frutescens
MH559326.1_Capsicum_frutescens
NC_033526.1_Capsicum_tovarii
KR078313.1_Capsicum_annuum_var._annuum
KJ619462.1_Capsicum_annuum_var._glabriusculum
KR078311.1_Capsicum_annuum_var._glabriusculum
NC_018552.1_Capsicum_annuum
MH559321.1_Capsicum_chinense
NC_030543.1_Capsicum_chinense
KX913217.1_Capsicum_chinense
MZ379791.1_Capsicum_chinense
NC_033524.1_Capsicum_galapagoense
NC_033527.1_Capsicum_eximium
KR078314.1_Capsicum_baccatum_var._baccatum
NC_039693.1_Capsicum_baccatum_var._baccatum
NC_039695.1_Capsicum_baccatum_var._praetermissum
NC_033525.1_Capsicum_chacoense
NC_039694.1_Capsicum_pubescens
NC_026551.1_Capsicum_lycianthoides
NC_062492.1_Lycianthes_radiata
AC_000188.1_Solanum_lycopersicum
CA Complex
CB Complex
Pubescens
Andean clade
Outgroup
-0.4
-0.71
-0.88
-0.5
-0.61
-1.22
-0.37
-0.8
-1.86
-0.16
-0.38
-0.99
-1.19
-1.75
-3.66
-5.83
-0.2
-0.47
-2.71
-8.72
-20.13
-25.79
-21.38
1
0.28
0.32
0.9
0.98
0.96
0.99
0.42
0.79
0.73
-25
-20
-15
-10
-5
0

Fig. 5 Phylogenetic analysis and estimation of divergence time of 22 *Capsicum* species based on maximum-likelihood analyses and Bayesian inference. *S. lycopersicum*, and *L. radiata* were used as the outgroup. The number mentioned on each node indicates estimated divergence time in My (Million years), and the number on each clade indicates Bayesian posterior probabilities from the BEAST analyses. Divergence time of *Capsicum* family was derived from TimeTree5 (http://www.timetree.org/) using the 'Get an evolutionary timeline' function

created using the Maximum Likelihood (ML) method with a bootstrap repeat of 1,000 under the Kimura 2-parameter model; *Solanum lycopersicum* and *Lycianthes radiata* were selected as the outgroup species. The ML phylogenetic analysis showed strong bootstrap support (between 55 and 100) for most branches, revealing five separate clades (data not shown). To determine evolution of *Capsicum* species, we repeated the ML phylogenetic analysis for divergence time estimation using the BEAST program, based on the uncorrelated relaxed lognormal clock model time using GTR+I+G category from the MAFFT alignment of all the 22 cp. genome sequences. The results of analysis of the MCMC tree related to species divergence times in *Capsicum* species are shown in Fig. 5. Speciation times were estimated in millions of years to assess the age of the clades.

Similar to the ML phylogenetic tree, the *C. tovarii*, accession of this study found with a *C. frutescens* accession along with *C. annuum* accessions with strong support in a different clade (Fig. 5). Similarly, *C. galapagoense* accession sequenced in this study was grouped with the *C. chinense* and *C. eximum* clades, and the *C. chacoense* accession sequenced in this study was shown to be a sister species strongly related to the *C. baccatum* complex. The divergence time estimation analysis verified that the *Capsicum* cp. genome diverged from the outgroups *S. lycopersicum* and *L. radiata* ~ 25.79 million years ago (Mya). The most recent common ancestor of all *Capsicum* cp. genome species was estimated to be ~ 20.13 Mya. As reported in an earlier study, the *C. pubescens* accession is a sister species to both the *C. annuum* and *C. baccatum* lineages, which are estimated to diverge at ~ 8.72 Mya. Our estimation of the *Capsicum* group speciation showed that the divergence time of *C. annuum* and *C. baccatum* complexes was about ~ 5.83 Mya. The cultivated varieties within *C. annuum* complexes such as, *C. annuum*, *C. frutescens*, *C. tovarii*, *C. chinense*, *C. eximum* and *C. galapagoense*. diverged approximately 3.66 Mya, while *C. eximum* and *C. galapagoense* subgroups diverged from *C. chinense* approximately 1.75 and 1.19 Mya respectively. Similarly, the domesticated *C. annuum* species started diverging ~ 1.86 Mya, while *C. frutescens* and *C. tovarii* subgroups diverged from *C. annuum* approximately 0.88 Mya. Thus, this study suggests that domesticated *C. annuum* and *C. chinense* are the most divergent species groups among the analyzed genotypes.

The PCA based on the entire cp. genome sequences showed five significant principal components (Fig. 6). The PCA results revealed that the first three principal components, PCA1, PCA2, and PCA3, accounted for 32.55%, 24.40%, and 30.76% of the total variation. In a manner consistent with the ML phylogenetic tree, the cultivated species within the *C. annuum* complex formed a closely related cluster. The accessions of *C. galapagoense* and *C. eximium* were grouped with *C. chinense*, whereas the accession of *C. chacoense* formed a closely related group with the *C. baccatum* complex.

Discussion

Typical characteristics of the cp. Genome

Plastid genomes are highly stable, non-recombinant, and uniparentally inherited with low mutation rates, making them a valuable source of information to produce highly reliable phylogenetic trees and to elucidate their evolutionary history [37, 38]. Though exception on different mode of inheritance and genome rearrangements have been reported in plastid genome. Advances in next-generation sequencing (NGS) techniques have significantly expedited and made sequencing of cp. genomes more cost-effective. In this study, we generated complete cp. genome sequences for three *Capsicum* species (Fig. 1A, B, and C) using NGS. We compared the

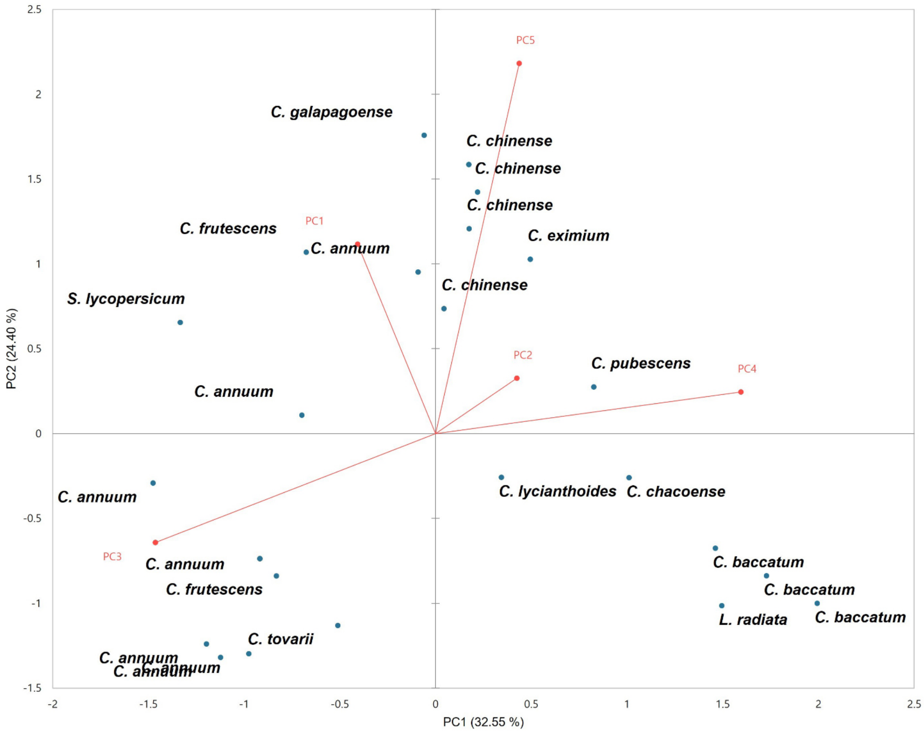

Fig. 6 Principal component analysis (PCA) of 22 *Capsicum* species based on cp. genome data. *S. lycopersicum* and *L. radiata* were used as outgroups

genomic features, content, and organization of the plastid genomes of 10 different *Capsicum* species (Table 1) that showed a typical quadripartite structure [39]. The length of the plastid genomes varied from 156,583 (*C. lycianthoides*) to 157,390 bp (*C.pubescens*), suggesting that among the Solanaceae family, variation in plastome size is a distinct feature of cultivated *Capsicum*, as reported previously [40]. The length of the LSC ranged from 86,813 bp (*C. lycianthoides*) to 87,688 bp (*C. pubescens*), whereas the SSC region was quite large (18, 522 bp) in the *C. lycianthoides* cp. genome. However, the length of the IR region was less variable, from 25,624 bp (*C.lycianthoides*) to 25,910 bp (*C. baccatum*). The GC content averaged around 37.72% in the 10 complete *Capsicum* cp. genomes, showing a strong similarity between them. Some instances of gene loss were identified in these cp. genomes (Table 1).

The variation observed in the cp. genome size was significantly affected by multiple factors, including gene loss, variations within IR regions, and differences in intergenic spacer regions, as reported in previous studies [41–43]. Within the set of the 10 *Capsicum* species investigated, 127–132 cp-coding genes were reliably identified. Notably, an absence of a single instance of the mRNA gene rps12 in the cp. genomes of *C. annuum*, *C. baccatum*, and *C. frutescens* was observed. Similarly, single copies of the mRNA genes *ycf1* and *ycf15* were absent from the cp. genome of *C. pubescens*. Of particular significance was the observation that there was a complete loss of the mRNA gene *psbZ* along with deletion of a copy of the *trnV-GAC* gene in the cp. genome of *C. lycianthoides*. Similarly, four mRNA genes, *psbB*, *rpl2*, *ycf1*, and *ycf15* were found to be lacking, whereas an additional mRNA gene, *lhbA*, was found to augment the cp. genome structure. Indeed, several notable changes in the cp. genomes of various plant species have been documented [44–46]. These include incorporation of the *psbZ* gene; absence of mRNA genes such as *lhbA*, *infA*, *rpl22*, and *rps16*; and instances of intron and copy deletions observed in genes such as *rpl2*, *clpP*, and *rps12*. These observations highlight the dynamic nature of evolution of the cp. genome across the plant kingdom.

Codon usage frequency analysis

Codon usage analysis is a fundamental tool for understanding the intricacies of genomic structure, evolutionary dynamics, and selective pressures imposed upon genes [47, 48]. The degeneracy inherent in the genetic code is manifested by its capacity to accommodate multiple codons to encode a singular amino acid. Codons that serve as interchangeable vehicles for a single amino acid are designated as synonymous codons. Regarding the three sequenced *Capsicum* species, the protein-coding genes showed a preference for codons ending in A or T/U. These codons had RSCU values of >1 (Supplementary Table S2). Similar findings were reported in the cp. genomes of a diverse array of flowering plants, an observation that can be attributed to the prevalent A/T-rich composition inherent to cp. genomes [49–52]. Similar codon usage patterns are evident in closely related plant species and those from the same lineage, providing insights into plant evolution. Species within the same genus show similar codon usage, even when compared to larger taxonomic groups such as families [53]. Therefore, studying codon usage is valuable for understanding evolutionary relationships within the *Capsicum* genus.

Evolution of cp. genome of *Capsicum* species

In flowering plants, recurrent contraction and expansion patterns at the junctions of SSC and LSC with IR regions have been identified as evolutionary indicators which serve to elucidate the relationships among different taxonomic groups [54]. Examination of boundaries within the cp. genome regions revealed distinct positional variations across the genomes of *Capsicum* species (Fig. 2). Higher variations observed at the extremities of single-copy and inverted repeats become conspicuous when profoundly diverged lineages are compared. In contrast, organisms that have undergone relatively recent evolutionary processes show reduced dissimilarity within vicinities of the boundaries [52, 55]. In the present study, the genus *Capsicum* exhibited a discernible pattern similar to this fluctuating motif at the junctures of single-copy and IR boundaries. This similarity could potentially signify recent evolutionary schisms among diverse *Capsicum* species.

Analysis of divergent hot spots

Analysis using mVISTA revealed a notable discrepancy in variation patterns, with the IR regions displaying reduced variability compared with the LSC and SSC regions. Moreover, the non-coding regions exhibited a higher number of variation sites than the coding regions (Fig. 3). This is in accordance with the observed diversity patterns prevalent in most angiosperms [56–58]. Among the analyzed genomes, some showed variations in the *ycf1* and *ycf2* genes, and intragenic regions of *accD*. Moreover, the results showed that there were more variable regions in the noncoding region than in the coding region of the *Capsicum* genome. The nucleotide diversity of the cp. genomes of the *Capsicum* species was approximately 0.00225 (Fig. 4). The *Capsicum* species showed a slightly lower average Pi value than that reported in a previous study [59]. In addition, four highly variable regions were identified within the cp. genomes of *Capsicum* species, making them potentially robust DNA barcodes. These regions encompassed intergenic regions (*trnT-UGU/trnL-UAA*) and four specific genes (*ccsA*, *ndhD*, *rps12*, and *ycf1*), all of which exhibited Pi values exceeding 0.01. A previous study reported that the *trnN–GUU/ndhF* intergenic region is a notable polymorphic hotspot, particularly when employing *C. lycianthoides* (NC_026551) as a reference genome [59]. Similarly, nucleotide polymorphism hotspots in cp. genomes were reported in other Solanaceae species [44, 60]. The genes with mutational sites, specifically *ccsA*, *ndhD*, *rps12*, and *ycf1*, are likely to play significant roles in facilitating adaptation of species to various environmental conditions [61]. For example, trans-splicing of plastid *rps12* transcripts is crucial for key embryonic development processes, such as cotyledon formation, provascular tissue establishment, and shoot apical meristem organization [62]. These mutational hotspots hold promise as potential loci for the development of novel DNA barcodes aimed to classify plants within the *Capsicum* genus.

Phylogenetic analysis

The cp. genome is highly conserved in sequence and structure, which helps determine molecular homology, consequently rendering it valuable in the construction of plant phylogenies [63]. In line with this, our investigation encompassed phylogenetic analysis of 22 *Capsicum* genotypes, including 10 distinct *Capsicum* species, with *S. lycopersicum* and *L. radiata* serving as designated outgroups. A dataset comprising 19 *Capsicum* cp. genomes was acquired from NCBI, and sequence alignment was performed using the MAFFT v7.475 alignment tool [30]. The evolutionary relationships among these chloroplast genomes revealed potential variation, and the categorization of individual species into clades was remarkably consistent with the findings of previous studies [7, 59]. However, the evolutionary relationships among *Capsicum* species have not been clearly clarified because of a lack of complete molecular data. Only a few cp. genomes have been published, and they do not include comprehensive species for genomic comparison and phylogenetic analyses [59, 64]. Hence, in this study, we selected three wild relatives in conjunction with domesticated *Capsicum* species to investigate cp. genome variations.

Cp genomic data may serve as an excellent source for highly accurate estimation of divergence time at the population level [60]. Hence, the divergence time for the 10 *Capsicum* species was estimated employing the GTR+I+G model of the BEAST2.0 program (Fig. 5). Estimation of the divergence time in the MCMCTree likelihood calculation with an uncorrelated relaxed lognormal clock model showed that the *Capsicum* species diverged from the outgroup species at ~25.79 million years ago (Mya). Using different time partitions, divergence time analysis considers variations in how rapidly genes evolve in different lineages [65]. This facilitates the acquisition of more accurate estimates of both the rates and ages of specific points in the evolutionary tree. *C. lycianthoides* was found to be the common ancestor of all *Capsicum* species, with an estimated speciation at ~20.13 Mya. The *C. lycianthoides* species was found in the Andean clade and was a sister clade to the rest of the genus as reported in previous studies [7, 66]. This study successfully divided all the *Capsicum* clades into major lineages, in alignment with previous reports [66], in which, among the scrutinized genotypes, the domesticated *Capsicum* species stood out as the most extensively divergent group. It has recently been reported that extensive domestication and selective breeding of *C. annuum* has led to a notably reduced level of nucleotide diversity in contrast to the other two domesticated species, *C. frutescens* and *C. chinense* [67].

In a recent time-calibrated phylogenetic analysis of the Solanaceae family, divergence of the *Capsicum* clade of both the *Solanum* and *Lycianthes* clades was estimated to have occurred at approximately 19 and 13 Mya, respectively [68]. However, significant speciation events within *Capsicum* likely occurred at approximately 1–3 Mya, contributing to contemporary species diversity within the genus. Domestication has led to modification of plant traits by influencing both the morphology and functionality of plants within agroecosystems [69]. Hence, it has been documented that cultivated plants exhibit a greater degree of trait variability than their wild progenitors [70–72]. The *Phaseolus* species (tepary beans) serves as a distinct model for investigating crop evolution; [73] analyzed SNP markers in both wild and domesticated *P. acutifolius* species and found that species of domesticated tepary beans clustered together, based on their origin. Similarly, [74] found a significant reduction in genetic diversity in lima beans (*P. lunatus*) owing to domestication. Several researchers have used different molecular markers and nucleotide data to show a reduction in the genetic diversity of various crop species compared to their wild progenitors [15, 75, 76].

Conservation and breeding implications for Pepper

Understanding the genetic variation in pepper species, particularly through cp. genome analysis, provides critical insights for both conservation and breeding efforts [25]. The domestication and intensive breeding of pepper over the past 5,000 years have significantly shaped its genetic landscape. According to the PCA and phylogenetic analysis, the cultivated species/varieties appeared as one structured monophyletic group. The cp. genome variation in pepper species offers valuable insights that are crucial for both conservation and breeding. By leveraging this knowledge, we can ensure the long-term sustainability and productivity of pepper crops, safeguarding them against future challenges while enhancing their value for both producers and consumers.

Conclusions

Our study effectively demonstrates the utility of cp-derived phylogenetic analysis in elucidating species relationships within taxonomically intricate and highly diverse groups, such as *Capsicum* species. Through comprehensive sequencing of complete cp. genomes (156–157 kbp), we achieved a high-resolution phylogenetic framework with robust support values. Our findings contribute significantly to the existing body of knowledge, emphasizing that the utilization of complete cp. genome sequences enhances phylogenetic resolution of *Capsicum* species, especially at the infrageneric taxonomic level. This improved resolution has unveiled narrowed levels of genetic variability within *Capsicum* species with possible genetic bottlenecks, as previously reported [11, 77]. Furthermore, it is evident that selective pressures exerted by early agriculturalists and contemporary crop-breeding practices have led to a notable reduction in the inherent genetic variability of cultivated crops. In conclusion, in the context of evolutionary analysis, it is observed that the early domesticated species of *C. annuum* exhibit a higher degree of divergence, very similar to their divergent phenotypes, when compared with their counterparts. These findings significantly broaden our understanding of *Capsicum* domestication and diversification and might play a pivotal role in advancing future research in the fields of evolution and phylogenetics related to *Capsicum* species.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05513-7.

Supplementary Material 1: **Supplementary Table S1**. The sample and the NCBI sequence information used in this study**Supplementary Table S2**. The relative synonymous codon usage in the C. chacoense, C. tovarii and C. galapagoense cp genomes

Author contributions

S.R. and J.K. Data curation, Investigation, Methodology, Software, and Writing the original draft. I-H.J. and J-K.Y. Experiments and analysis. W.J., S.H. and H-S.P. Formal analysis, Visualization, and Validation. A.M.AG. and A.A.H. Contributed additional data. D.S. and Y-S.S. Data curation, Investigation, and Methodology. J-W.C. Conceptualization, Supervision, Funding, and Project administration. All authors read and approved the final version of the manuscript.

Funding

This work was carried out with the support of "Co-operative Research Program for Agriculture Science and Technology Development (Project No. PJ017066)" Rural Development Administration, Republic of Korea. The authors extend their appreciation to the Researchers supporting project number (RSPD2024R931) King Saud University, Riyadh, Saudi Arabia.

Data availability

All sequences and annotation data were deposited in a public database at the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) under accession numbers NC_033525, NC_033524, and KX913219.

Declarations

Ethics approval and consent to participate

The authors confirm that all methods comply with local and national regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 May 2024 / Accepted: 12 August 2024
Published online: 24 August 2024

References

1. Mohd Hassan N, Yusof NA, Yahaya AF, Mohd Rozali NN, Othman R. Carotenoids of Capsicum fruits: Pigment Profile and Health-promoting functional attributes. Antioxid (Basel) 2019, 8(10).
2. Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC. Genetic diversity and structure in semiwild and domesticated chiles (Capsicum annuum; Solanaceae) from Mexico. Am J Bot. 2009;96(6):1190–202.
3. De AK. Capsicum: the genus Capsicum. CRC; 2003.
4. Luo XJ, Peng J, Li YJ. Recent advances in the study on capsaicinoids and capsinoids. Eur J Pharmacol. 2011;650(1):1–7.
5. Hayman M, Kam PCA. Capsaicin: a review of its pharmacology and clinical applications. Curr Anaesth Crit Care. 2008;19(5):338–43.
6. Barboza GE, García CC, Bianchetti LdB, Romero MV, Scaldaferro M. Monograph of wild and cultivated Chili peppers (Capsicum L., Solanaceae). PhytoKeys. 2022;200:1–423.
7. Carrizo García C, Barfuss MH, Sehr EM, Barboza GE, Samuel R, Moscone EA, Ehrendorfer F. Phylogenetic relationships, diversification and expansion of Chili peppers (Capsicum, Solanaceae). Ann Bot. 2016;118(1):35–51.
8. Barboza GE, Carrizo García C, Leiva González S, Scaldaferro M, Reyes X. Four new species of Capsicum (Solanaceae) from the tropical Andes and an update on the phylogeny of the genus. PLoS ONE. 2019;14(1):e0209792.
9. Barboza GE, García CC, Scaldaferro M, Bohs L. An amazing new Capsicum (Solanaceae) species from the andean-amazonian Piedmont. PhytoKeys. 2020;167:13–29.
10. Perry L, Dickau R, Zarrillo S, Holst I, Pearsall DM, Piperno DR, Berman MJ, Cooke RG, Rademaker K, Ranere AJ, et al. Starch fossils and the domestication and Dispersal of Chili Peppers (Capsicum Spp. L.) in the Americas. Science. 2007;315(5814):986–8.
11. Tripodi P, Rabanus-Wallace MT, Barchi L, Kale S, Esposito S, Acquadro A, Schafleitner R, van Zonneveld M, Prohens J, Diez MJ, et al. Global range expansion history of pepper (Capsicum spp.) revealed by over 10,000 genebank accessions. Proc Natl Acad Sci. 2021;118(34):e2104315118.
12. Zou X, Ma Y, Dai X, Li X, Yang S. Spread and Industry Development of Pepper in China. Acta Horticulturae Sinica. 2020;47(9):1715–6.
13. Paran I, van der Knaap E. Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. J Exp Bot. 2007;58(14):3841–52.
14. Bosland PW, Votava EJ. Peppers: Vegetable and Spice Capsicums. CABI; 2012.
15. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol. 2012;30(1):105–11.
16. Cao Y, Zhang K, Yu H, Chen S, Xu D, Zhao H, Zhang Z, Yang Y, Gu X, Liu X, et al. Pepper variome reveals the history and key loci associated with fruit domestication and diversification. Mol Plant. 2022;15(11):1744–58.
17. Peery mz. Kirby r, reid bn, stoelting r, doucet-bëer e, robinson s, vásquez-carrillo c, pauli jn, palsbøll pj: reliability of genetic bottleneck tests for detecting recent population declines. Mol Ecol. 2012;21(14):3403–18.
18. Martins KC, Pereira TNS, Souza SAM, Rodrigues R. Amaral Junior ATd: crossability and evaluation of incompatibility barriers in crosses between < b > Capsicum species. Crop Breed Appl Biotechnol 2015, 15.
19. Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG. Metabolite biodiversity in pepper (Capsicum) fruits of thirty-two diverse accessions: variation in health-related compounds and implications for breeding. Phytochemistry. 2011;72(11–12):1358–70.

20. Sathiyamurthy VA, Veeraragavathatham D, Chezhiyan N. Studies on the capsaicin content in Chilli hybrids. Capsicum Eggplant Newsl. 2002;21:44–7.
21. Hyun DY, Sebastin R, Lee G-A, Lee KJ, Kim S-H, Yoo E, Lee S, Kang M-J, Lee SB, Jang I, et al. Genome-wide SNP markers for genotypic and phenotypic differentiation of Melon (Cucumis melo L.) varieties using genotyping-by-sequencing. Int J Mol Sci. 2021;22(13):6722.
22. Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, Seo E, Choi J, Cheong K, Kim K-T, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet. 2014;46(3):270–8.
23. Barchi L, Pietrella M, Venturini L, Minio A, Toppino L, Acquadro A, Andolfo G, Aprea G, Avanzato C, Bassolino L, et al. A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Sci Rep. 2019;9(1):11769.
24. Liao Y, Wang J, Zhu Z, Liu Y, Chen J, Zhou Y, Liu F, Lei J, Gaut BS, Cao B, et al. The 3D architecture of the pepper genome and its relationship to function and evolution. Nat Commun. 2022;13(1):3479.
25. Magdy M, Ou L, Yu H, Chen R, Zhou Y, Hassan H, Feng B, Taitano N, van der Knaap E, Zou X, et al. Pan-plastome approach empowers the assessment of genetic variation in cultivated Capsicum species. Hortic Res. 2019;6(1):108.
26. Wambugu PW, Brozynska M, Furtado A, Waters DL, Henry RJ. Relationships of wild and domesticated rices (Oryza AA genome species) based upon whole chloroplast genome sequences. Sci Rep. 2015;5(1):13957.
27. Wright F. The 'effective number of codons' used in a gene. Gene. 1990;87(1):23–9.
28. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14(13):5125–43.
29. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003;13(4):721–31.
30. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017;20(4):1160–6.
31. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
32. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
33. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29(8):1969–73.
34. Yule GU, Willis FRS. Philosophical Trans Royal Soc Lond Ser B Containing Papers Biol Character. 1925;213(402–410):21–87.
35. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67(5):901–4.
36. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5.
37. Palmer JD, Shields CR, Cohen DB, Orton TJ. Chloroplast DNA evolution and the origin of amphidiploid Brassica species. Theor Appl Genet. 1983;65(3):181–9.
38. Nikiforova SV, Cavalieri D, Velasco R, Goremykin V. Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. Mol Biol Evol. 2013;30(8):1751–60.
39. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3–5):273–97.
40. Jo YD, Park J, Kim J, Song W, Hur CG, Lee YH, Kang BC. Complete sequencing and comparative analyses of the pepper (Capsicum annuum L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. Plant Cell Rep. 2011;30(2):217–29.
41. Lei WJ, Ni DP, Wang YJ, Shao JJ, Wang XC, Yang D, Wang JS, Chen HM, Liu C. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus Membranaceus. Sci Rep 2016, 6.
42. Zhang YJ, Du LW, Liu A, Chen JJ, Wu L, Hu WM, Zhang W, Kim K, Lee SC, Yang TJ et al. The Complete Chloroplast Genome Sequences of Five Epimedium Species: lights into phylogenetic and taxonomic analyses. Front Plant Sci 2016, 7.
43. Amiryousefi A, Hyvonen J, Poczai P. The chloroplast genome sequence of bittersweet (Solanum dulcamara): plastid genome structure evolution in Solanaceae. PLoS ONE 2018, 13(4).
44. Feng SG, Zheng KX, Jiao KL, Cai YC, Chen CL, Mao YY, Wang LY, Zhan XR, Ying QC, Wang HZ. Complete chloroplast genomes of four Physalis species (Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships. Bmc Plant Biol 2020, 20(1).
45. Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. Complete plastid genome sequence of the chickpea (Cicer arietinum) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). Mol Phylogenet Evol. 2008;48(3):1204–17.
46. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Giray JC, Morden CW, et al. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 2001;13(3):645–58.
47. Morton BR. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. J Mol Evol. 1998;46(4):449–59.
48. Goodarzi H, Torabi N, Najafabadi HS, Archetti M. Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons. Gene. 2008;407(1):30–41.
49. Chen XC, Li QS, Li Y, Qian J, Han JP. Chloroplast genome of Aconitum barbatum var. Puberulum (Ranunculaceae) derived from CCS reads using the PacBio RS platform. Front Plant Sci 2015, 6.
50. Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA, Lee SM, Lee IJ. Complete chloroplast genome of Nicotiana otophora and its comparison with related species. Front Plant Sci 2016, 7.
51. Abdullah WS, Mirza B, Ahmed I, Waheed MT. Comparative analyses of chloroplast genomes of Theobroma cacao and Theobroma grandiflorum. Biologia. 2020;75(5):761–71.
52. Abdullah SI, Mehmood F, Ali Z, Malik MS, Waseem S, Mirza B, Ahmed I, Waheed MT. Comparative analyses of chloroplast genomes among three Firmiana species: identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene. 2019;19:100199.
53. Abdullah MF, Rahim A, Heidari P, Ahmed I, Poczai P. Comparative plastome analysis of Blumea, with implications for genome evolution and phylogeny of Asteroideae. Ecol Evol. 2021;11(12):7810–26.
54. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. Bmc Evol Biol 2008, 8.
55. Shahzadi I, Abdullah, Mehmood F, Ali Z, Ahmed I, Mirza B. Chloroplast genome sequences of Artemisia maritima and Artemisia absinthium: comparative analyses, mutational hotspots in genus Artemisia and phylogeny in family Asteraceae. Genomics. 2020;112(2):1454–63.
56. Lee S-R, Kim K, Lee B-Y, Lim CE. Complete chloroplast genomes of all six Hosta species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. BMC Genomics. 2019;20(1):833.
57. Yu J, Fu J, Fang Y, Xiang J, Dong H. Complete chloroplast genomes of Rubus species (Rosaceae) and comparative analysis within the genus. BMC Genomics. 2022;23(1):32.
58. Jeon JH, Kim SC. Comparative analysis of the complete chloroplast genome sequences of three closely related East-Asian Wild roses (Rosa sect. Synstylae; Rosaceae). Genes (Basel) 2019, 10(1).
59. D'Agostino N, Tamburino R, Cantarella C, De Carluccio V, Sannino L, Cozzolino S, Cardi T, Scotti N. The complete plastome sequences of Eleven Capsicum genotypes: insights into DNA variation and molecular evolution. Genes (Basel) 2018, 9(10).
60. Wang SB, Gao JP, Chao HY, Li ZW, Pu WX, Wang YF, Chen M. Comparative chloroplast genomes of Nicotiana Species (Solanaceae): insights into the genetic variation, phylogenetic relationship, and Polyploid Speciation. Front Plant Sci 2022, 13.
61. Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, Li ZH. Molecular evolution of Chloroplast genomes of Orchid species: insights into phylogenetic relationship and adaptive evolution. Int J Mol Sci 2018, 19(3).
62. Tadini L, Ferrari R, Lehniger MK, Mizzotti C, Moratti F, Resentini F, Colombo M, Costa A, Masiero S, Pesaresi P. Trans-splicing of plastid rps12 transcripts, mediated by AtPPR4, is essential for embryo patterning in Arabidopsis thaliana. Planta. 2018;248(1):257–65.
63. Yang JP, Zhang FW, Ge YJ, Yu WH, Xue QQ, Wang MT, Wang HM, Xue QY, Liu W, Niu ZT et al. Effects of geographic isolation on the Bulbophyllum chloroplast genomes. Bmc Plant Biol 2022, 22(1).
64. Shiragaki K, Yokoi S, Tezuka T. Phylogenetic Analysis and molecular diversity of Capsicum based on rDNA-ITS Region. Horticulturae 2020, 6(4).
65. Duchêne S, Ho SYW. Using multiple relaxed-clock models to estimate evolutionary timescales from DNA sequence data. Mol Phylogenet Evol. 2014;77:65–70.

66. Carrizo García C, Barboza GE, Palombo N, Weiss-Schneeweiss H. Diversification of chiles (Capsicum, Solanaceae) through time and space: new insights from genome-wide RAD-seq data. Front Genet 2022, 13.
67. Liu F, Zhao J, Sun H, Xiong C, Sun X, Wang X, Wang Z, Jarret R, Wang J, Tang B, et al. Genomes of cultivated and wild Capsicum species provide insights into pepper domestication and population differentiation. Nat Commun. 2023;14(1):5487.
68. Särkinen T, Bohs L, Olmstead RG, Knapp S. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. Bmc Evol Biol. 2013;13:214.
69. Meyer RS, DuVal AE, Jensen HR. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytol. 2012;196(1):29–48.
70. Robinson ML, Schilmiller AL, Wetzel WC. A domestic plant differs from its wild relative along multiple axes of within-plant trait variability and diversity. Ecol Evol. 2022;12(1):e8545.
71. Chacón-Labella J, García Palacios P, Matesanz S, Schöb C, Milla R. Plant domestication disrupts biodiversity effects across major crop types. Ecol Lett. 2019;22(9):1472–82.
72. Milla R, Bastida JM, Turcotte MM, Jones G, Violle C, Osborne CP, Chacón-Labella J, Sosinski ÊE Jr., Kattge J, Laughlin DC, et al. Phylogenetic patterns and phenotypic profiles of the species of plants and mammals farmed for food. Nat Ecol Evol. 2018;2(11):1808–17.
73. Gujaria-Verma N, Ramsay L, Sharpe AG, Sanderson L-A, Debouck DG, Tar'an B, Bett KE. Gene-based SNP discovery in tepary bean (Phaseolus acutifolius) and common bean (P. Vulgaris) for diversity analysis and comparative mapping. BMC Genomics. 2016;17(1):239.
74. Motta-Aldana JR, Serrano-Serrano ML, Hernández-Torres J, Castillo-Villamizar G, Debouck DG, Chacóns MI. Multiple origins of Lima Bean Landraces in the Americas: evidence from Chloroplast and Nuclear DNA polymorphisms. Crop Sci. 2010;50(5):1773–87.
75. Glémin S, Bataillon T. A comparative view of the evolution of grasses under domestication. New Phytol. 2009;183(2):273–90.
76. Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, Li M-W, He W, Qin N, Wang B, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet. 2010;42(12):1053–9.
77. Lozada DN, Bhatta M, Coon D, Bosland PW. Single nucleotide polymorphisms reveal genetic diversity in New Mexican Chile peppers (Capsicum spp). BMC Genomics. 2021;22(1):356.

Publisher's Note