BMC Genomics

RESEARCH Open Access

Mitochondrial genome characterization and new insights on phylogenetic relationship of *Chara* (Charophyceae, Charophyta)

Yuhui Xie[1,2], Hanlin Bai[1,2], Xudong Liu[1,2], Jia Feng[1,2], Shulian Xie[1,2*] and Fangru Nan[1,2*]

Abstract

Background Charophyta, initially classified as ferns and later within green algae, were recognized as an independent division in the twentieth century. The taxonomic status of Charophyta has evolved considerably over time and phylogenomic studies elucidating the evolutionary relationships within Charophyta remain limited. This study sequenced and analyzed the mitochondrial genomes of eight charophyte plants, including genome size, gene content, and codon usage bias, and compared these features with those of other algal lineages and higher plants.

Results Total DNA was extracted from the plant samples and subjected to single-molecule sequencing and genomic sequencing, respectively. The successfully sequenced *rbc*L and genomic sequences were used to reconstruct the phylogenetic tree. The single-gene phylogeny indicated that the sample sxuxm2407072 (xm2) is a close relative of *Chara canescens* Loiseleur, while the remaining samples are affiliated with *Chara vulgaris* Linn. Additionally, xm2 exhibited distinct spine cell morphology compared to other samples. Eight mitochondrial genomes were assembled in this study, ranging from 66,695 to 70,447 bp in size and exhibiting slightly lower GC content than AT content, yet a significantly greater GC skew compared to the AT skew. RSCU analysis revealed consistent codon usage patterns across samples, with AGA representing the most frequently used codon. Analysis of codon usage bias in the eight mitochondrial genomes yielded values for CAI, CBI, GC12, GC3s, and ENC. The results showed extremely low CAI, consistently negative CBI, a lack of correlation between GC12 and GC3s, and a significant positive correlation between ENC and GC3s. Phylogenetic reconstruction based on mitochondrial genomic data consistently revealed eight plant samples as a sister group to the *C. vulgaris* clade and highly congruent topological arrangements.

Conclusions The newly assembled eight *Chara* mitochondrial genomes in our study revealed complex structures and relatively conserved sequences, exhibiting limited divergence across the entire genome while displaying elevated variability in specific gene regions. The mitochondrial genomes underwent nearly neutral evolution in their codon usage patterns. Based on mitochondrial genome phylogenetic analysis, the mitochondrial genomes of *Chara* plants in this study showed consistent phylogenetic signals, indicating a similar evolutionary pathway. The mitochondrial

*Correspondence:
Shulian Xie
xiesl@sxu.edu.cn
Fangru Nan
nanfr@sxu.edu.cn

Full list of author information is available at the end of the article

data generated in this study provide new insights into the evolutionary history of charophytes, enrich the genomic resources for this group, and establish a solid foundation for future research on this important lineage.

Keywords Charophyta, Mitochondrion, Codon usage bias, Phylogenomics

Background

Charophyta is the most morphologically complex group in terms of morphology among green algal groups and is considered a crucial step in the terrestrialization of plants [1, 2]. The morphological structures of this lineage are analogous to those of land plants, characterized by rhizoids extending into the substrate and stems exhibiting green to gray-green coloration, often encrusted with substantial calcium deposits. At the stem nodes, branchlets develop which bear orange-red to golden-brown gametangia, which consist of oogonia and antheridia. Charophyta exhibit a broad geographical distribution, inhabiting diverse wetland, freshwater, and brackish environments across tropical, temperate, and polar regions [3, 4]. Members of the genus *Chara* (Charophyta) display considerable plasticity across a broad range of morphological traits and habitat preferences. *C. vulgaris* exhibits pronounced morphological variation, which is evident even among individuals within a single population or clonal strain [5]. However, the species has experienced both habitat contraction and population loss, largely resulting from anthropogenic pressures including water pollution and habitat destruction [3]. Current evidence supports a streptophyte algal origin for plant terrestrialization, whereby early charophytes acquired a suite of genetic, metabolic, and morphological adaptations that facilitated their colonization of terrestrial environments. For half a century, studies of algae have highlighted the evolutionary significance of *Chara*, which has served as a key model system for insights into plant cell biology, development, physiology, and ecology [6].

Discussions regarding the taxonomic status of Charophyta have always existed. In early studies, phylogenetic investigations of Charophyta were primarily based on morphological characteristics. Taxonomists rely on morphological distinctions in stipulodes, cortical patterns, branchlet architecture, and reproductive structures as diagnostic features for species identification. In 1965, Wood & Imahori proposed phylogenetic hypotheses based on morphological characters for extant genera of the Characeae [7]. Khan & Sarma (1984) also proposed a corresponding phylogenetic hypothesis based on both morphological features and chromosome numbers, suggesting that Characeae is a monophyletic lineage [8]. Molecular phylogenies have confirmed that not only are the individual charophyte lineages monophyletic, but also that they, together with land plants, form a monophyletic group [9–12]. In the latter half of the twentieth century, Charophyta was classified within the division Chlorophyta, based on ultrastructural characteristics of cell morphology and division [10, 13]. The rapid development of sequencing technology has established molecular phylogenetics as a primary tool for evolutionary study [14]. However, the most significant conflict between the 18S and *rbc*L analyses concerned the placement of the Charales, thereby precluding a definitive determination of their taxonomic status [15]. By analyzing plastid (*atp*B, *rbc*L), mitochondrial (*nad*5), and nuclear (SSU rRNA) gene sequences, Karol and Straub identified Charales as the closest living relatives of land plants and recovered Coleochaetales as sister to this Charales-land plant clade [13, 14]. In current phylogenetic frameworks, Charophyceae, Coleochaetophyceae, and Zygnematophyceae are classified within the clade Phragmoplastophyta, which comprises these three algal lineages and the land plants [2].

Organelle genomics has provided powerful insights for elucidating algal evolutionary history due to technological advancements. The advancement of next-generation sequencing technologies has enabled the acquisition of more extensive sequence data, encompassing both organelle genomes and nuclear gene sequences [16]. The advent of high-throughput sequencing technology facilitates the reconstruction of phylogenetic relationships with unprecedented precision, both within and across species boundaries. Comparative mitochondrial genomics further reveals inter-individual diversity in terms of gene structure, content, and composition [17]. Comprehensive analysis of published mitochondrial genomic data reveals similar evolutionary patterns across diverse lineages. A predominant feature is the 'circular-mapping' architecture, which typically corresponds to linear, head-to-tail concatemeric DNA molecules [18, 19]. During the evolution of Streptophyhta, the structure of mitochondrial DNA has undergone substantial changes, including the insertion of introns, gene loss, gene order shuffling, and mitochondrial genome amplification [20]. The organelle genome data of *Chara* is limited, with only *C. vulgaris* and *Chara braunii* C.C.Gmelin having relevant data released. The related researches provide the organelle genome sequence of *C. vulgaris* and compare the contents and types of introns between *C. vulgaris* with several other species (including land plants). Although the organelle genomes of *C. braunii* have been assembled from whole-genome sequencing data, they remain largely unannotated and inadequately analyzed [2, 21].

This study aimed to characterize the mitochondrial genomes of eight plants, elucidate their codon usage

bias patterns, and infer phylogenetic relationships using genomic data. The taxonomic status of the eight samples was determined using *rbc*L sequences and morphological characteristics. Following the sequencing and assembly of their mitochondrial genomes, the mitochondrial genomic features were compared among individuals. Phylogenetic relationships were subsequently reconstructed using Bayesian inference and maximum likelihood methods.

Materials and methods

Sample collecting

Samples were collected from multiple locations in China as in Table 1. Freshly collected algal material was subjected to a rigorous cleaning procedure. It was rinsed thoroughly with sterile distilled water (10–15 repetitions) until all visible debris was removed, followed by a 1–2 min immersion in Phosphate Buffer Saline (PBS) buffer. Upon microscopic examination, the cleaned algae are considered qualified samples if they are devoid of attached contaminating species such as diatoms and green algae. Otherwise, the aforementioned procedure is repeated. Excess moisture was subsequently removed by blotting with a sterile lint-free cloth.

Morphological observation

The cleaned specimens were immediately mounted on glass slides for microscopic observation. Observations were performed using an Olympus BX51 light microscope.

Sequence amplification

Genomic DNA extraction was conducted using approximately 0.5 g silica gel dried material. The material was thoroughly pulverized in liquid nitrogen with a mortar and pestle, and the DNA was purified with the Easy Pure Plant Genomic DNA Kit (Quanshijin Biotechnology Co., Ltd., Beijing, China) in accordance with the manufacturer's protocol. The optimal PCR primer sequences and amplification system for the *rbc*L gene of the Characeae were according to Song et al. (2022) [22]. After amplifying the PCR products, they were checked via 1% agarose gel electrophoresis (120 V, 20 min). The amplification products were sent to BGI for bidirectional sequencing by Sanger dideoxy termination method (BGI, Beijing).

Table 1 Sample collection information statistics

Strains ID	Collection location	Collection time	Specimen collector
sxupm2407041 (pm1)	Lvliang, Shanxi, China	July 4, 2024	Yang Liu
sxupm2407042 (pm2)	Lvliang, Shanxi, China	July 4, 2024	Yang Liu
sxuzm2406131 (zm1)	Taiyuan, Shanxi, China	June 13, 2024	Kehui Li, Tao Wang
sxuzm2406132 (zm2)	Taiyuan, Shanxi, China	June 13, 2024	Kehui Li, Tao Wang
sxuxm2407072 (xm2)	Lhasa, Tibet, China	July 4, 2024	Weinan Guo
sxuxm2407073 (xm3)	Lhasa, Tibet, China	July 7, 2024	Weinan Guo
sxuSH241014 (SH)	Shanghai, China	October 14, 2024	Shulian Xie
sxuXJ2410111 (XJ1)	Urumqi, Xinjiang, China	October 11, 2024	Weinan Guo

Genome sequencing, assembling and annotation

Whole-genome sequencing of the cleaned algal material was conducted by Novogene (Tianjin, China) using an Illumina high-throughput platform. The service was configured for 1 × coverage of a two Gb genome, producing a raw output of two Gb. DNA samples that passed quality control on Agilent 5400 were subsequently used for library preparation. A summary of the quality metrics from the DNA sequencing is available in Supplementary Material 1. Quality control was performed on the raw data to obtain clean data. Raw reads were subjected to quality control using the fastq software with the following parameter settings: qualified quality phred: 5, unqualified percent limit: 50, length required: 150, overlap diff limit: 1, and overlap diff percent limit: 10. The mitochondrial genome was extracted and assembled from the whole-genome sequencing data using GetOrganelle v1.7 + [23]. The GetOrganelle software was run using the -s parameter to specify the reference genome as the seed sequence, with K-mer values set to 21, 45, 65, 85, and 105, and the number of iterative cycles set to 20. The accuracy of the assembly was verified through BLAST analysis [24]. We primarily focused on the metrics of query score, percent identity, and e-value. Sequences exhibiting a query score and percent identity of 99–100% were defined as consistent and retained as correct assembly results. Assembly results failing to meet these thresholds were subsequently removed.

Based on a sequence identity criterion of 99–100%, correct assemblies were retained, and those failing to meet this threshold were removed. The Bandage software was used to visualize the circular mitochondrial genomes [25]. If contigs were not assembled into a complete circular structure, manual correction was performed using Bandage software. The resulting circular data was exported and blasted on NCBI to verify if it is a mitochondrial genome of Charophyta. In this study, the assembly and annotation of the mitochondrial genome sequence were completed with reference to the GenBank archived sequence (reference sequence accession number: ON406421) [26]. To identify open reading frames (ORFs) within the mitochondrial genome, the Ugene software was employed using the standard genetic code

[27]. The longest putative ORF sequences were subsequently subjected to a BLAST search against the reference genome to determine the optimal ORF region. The region with a matching rate of 99%–100% is the optimal ORF and is mapped to the reference genome. The final CDS region is determined by identifying the optimal ORF region and comparing it with the reference group. The amino acid sequence translated from the segment was verified using Sequin software (http://www.ncbi.nlm.nih.gov/Sequin). The tRNA genes were predicted using the ARAGORN online platform [28], whereas rRNA sequences were identified by alignment with known rRNA loci in the reference genome. Subsequent manual annotation of rRNA and tRNA features was performed separately using Sequin to generate structured annotation files. Finally, these annotation files were visualized with OGDRAW to produce a circular mitochondrial genome map [29]. The obtained annotation file underwent rigorous validation using GB2sequin and the verified files were submitted to NCBI database [30].

Phylogenetic analyses

*Rbc*L sequence of *Chara* and other related sequences from NCBI were selected to construct phylogenetic trees by using maximum likelihood (ML) and bayesian inference (BI) methods. The two species from Coleochaetophyceae selected as outgroups were *Coleochaete orbicularis* Pringsheim and *Coleochaete nitellarum* Jost according to previous literature [31].

We compiled a core dataset for phylogenetic analysis by extracting the available mitochondrial protein-coding genes, RNA sequences, and their corresponding amino acid sequences from all taxonomic groups. The mitochondrion genomes of *C. orbicularis* and *C. nitellarum* were obtained from GenBank. MAFFT v7, MACSE v2, Gblocks, and trimAl were used to optimize and trim protein-coding gene sequences, RNA sequences, and amino acid sequences. Poorly aligned regions for each individual gene alignment were then identified and removed [32–35]. The extracted sequences were concatenated to construct a gene tree using the concatenation method. Given that the samples in this study do not span multiple species and the mitochondrial gene sequences are relatively conserved, concatenated-based method was selected for phylogenetic analysis. The optimal partitioning scheme and corresponding evolutionary models were determined for each dataset using PartitionFinder2 [36]. The optimal evolutionary model obtained was automatically imported into the subsequent tree-building workflow. Maximum likelihood analysis was conducted using iqtree v2.4.0 on the concatenated dataset partitioned by gene [37], with 1000 rapid bootstrap replicates. Bayesian inference analysis of the gene-partitioned dataset was performed in MrBayes v3.2.6 with one cold and three heated Markov chains [38]. The analysis was run for 1 million generations with sampling conducted every 1000 generations, resulting in 1000 sampled trees. The first 25% of samples were discarded as burn-in, which represents the initial, non-stationary phase of the Markov chain, convergence was assessed using the average standard deviation of split frequencies (ASDSF), with a threshold value of 0.01 [20]. The orthologous genomes and single-copy orthologous genes of 12 species were analyzed using OrthoFinder [39]. The species phylogenetic tree based on orthologous genes was obtained using the species tree inference from all genes (STAG) method [40].

Codon bias analysis

Codon usage analysis was performed on 12 mitochondrial genes using codonW software (v1.4.4, http://codonw.sourceforge.net). Parameters including the effective number of codons (ENC), relative synonymous codon usage (RSCU), codon bias index (CBI), codon adaptation index (CAI), and nucleotide composition (GC, GC12 and GC3s) were calculated. A data fit was performed between GC12 and GC3s values. The resulting data were visualized and subjected to further statistical analysis using R. A comprehensive set of metrics derived from the codon usage bias analysis is presented in Supplementary Material 2.

The Effective Number of Codons (ENC) is a measure of codon usage bias, reflecting the degree of deviation from random synonymous codon selection. Lower ENC values indicate stronger bias. An ENC-plot graph is drawn with GC3 of the mitochondrial genome as the x-axis and ENC as the y-axis. The standard curve calculation formula is $y = 2 + x + 29/(x^2 + (1—x)^2)$ [41]. To quantify the deviation of the sampled points, we calculated the vertical distance between each point and the standard curve. Additionally, a pearson correlation analysis was conducted to assess the relationship between ENC and GC3. Detailed results are provided in the Supplementary Material 3.

Results

Analysis of single-gene phylogenetic tree

Two phylogenetic methods were used to analyze the *rbc*L gene, with *C. orbicularis* and *C. nitellarum* from Coleochaetophyceae as outgroups. Phylogenetic analyses revealed a consistent topology between ML and Bayesian methods (Fig. 1a), indicating that *Chara* is a paraphyletic group. Within the paraphyletic *Chara* clade, three distinct lineages were identified: one comprising zm2 and XJ1, a second containing zm1 and pm1, and a third consisting of xm3 and pm2. The three lineages formed a sister group to *C. vulgaris*, albeit with relatively weak bootstrap values. In contrast, sample SH formed a highly supported separate clade (posterior probability,

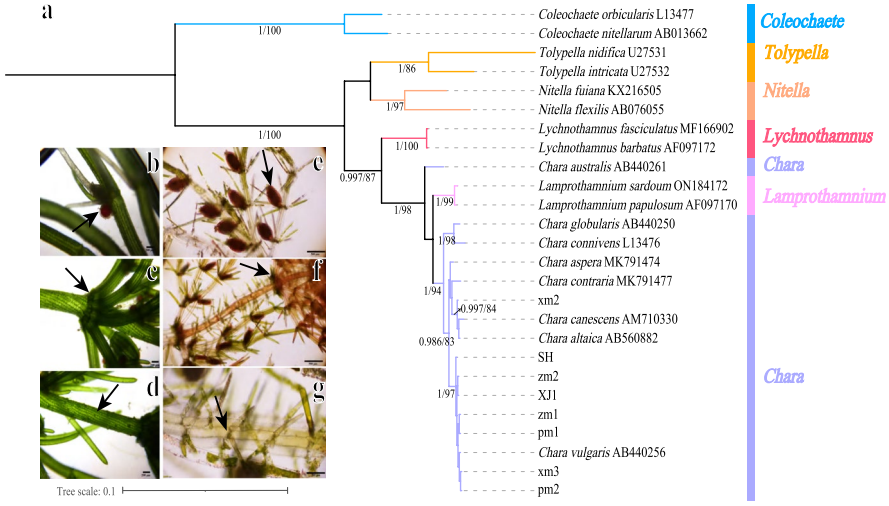

Fig. 1 Morphological characteristics of *Chara* and phylogenetic tree based on *rbc*L gene sequence (Support values for all analyses are noted as: Bayesian posterior probabilities/ML supporting values). **a** The phylogenetic tree constructed from *rbc*L sequences. **b**, **c** Antheridium and Oogonium (black arrow). **d**, **e** whorl with stipulodes (black arrow). **f**, **g** spine cells (black arrow)

PP = 1.00, Bootstrap = 97) and did not cluster with any other *Chara* samples. The phylogenetic analysis placed sample xm2 as sister to a clade comprising *Chara altaica* A. Braun and *C. canescens* (posterior probability, PP = 0.997, Bootstrap = 84), clearly separating it from *C. vulgaris*.

Morphological observation

Detailed microscopic examination of all eight specimens revealed key morphological characteristics of their stipulodes, cortex, reproductive organs, and apical branches (pm1, pm2, SH, xm3, zm1, zm2, XJ1, and xm2). The first seven accessions (pm1 to XJ1) exhibited highly similar morphologies. They were consistently monoecious, with male and female gametangia intermixed on the lower 2–4 nodes of the branches (Fig. 1b). The stipules were double whorled, with both whorls of similar length, robust, and blunt at the tip (Fig. 1c). The spine cells were solitary and papillate, and the stem had a two-layered cortex (Fig. 1d). For the xm3 plant, the reproductive organs were not observed due to their detachment. Some or all of the branches in the whorl lacked cortex. In the zm2 plant, the branches and inner bracts were longer. The eighth xm2 plant was dioecious, and this one was a female plant, with a single ovary (Fig. 1e). All bract cells were well-developed, acuminate, and the stipules were double whorled, with both upper and lower whorls well-developed and acuminate (Fig. 1f). The spine cells were either solitary or clustered, acuminate, and the stem had a regular single-layered cortex (Fig. 1g), with the top 1–2 node segments lacking cortex.

Assembly and main features of mitochondrial genome

The mitochondrial genomes of 8 specimens collected from the wild were obtained through high-throughput sequencing. Mitochondrial genomes were selected as the focus to address a known data gap for charophytes and to enhance genetic diversity resources. Compared with the chloroplast genomes, the mitochondrial genomes in our samples demonstrated superior assembly quality. This superior quality is essential for ensuring the reliability of subsequent analyses. The main features of the mitochondrial genomes of the 8 plants with those of *Nitellopsis obtusa* (MW556320), *Nitella hyalina* (NC_017598), and *Chara vulgaris* (ON406421, NC_005255) were compared as shown in Table 2. The mitochondrial genomes of *Nitellopsis obtusa* (MW556320), *Nitella hyalina* (NC_017598), and *Chara vulgaris* (ON406421, NC_005255) represent all published and fully annotated Characeae mitochondrial data available in the NCBI nucleotide database. Their inclusion in our analysis thus encompasses the existing taxonomic breadth within this family. The largest mitochondrial genome is from plant zm2, with a sequence length of 70,447 bp, while the smallest is from plant XJ1, with a sequence length of 66,695 bp. The GC content for the mitochondrial genomes is 42.1%, 41.8%, 42.1%, 42.2%, 42.2%, 42.2%, 41.9%, and 41.8%, respectively, with no significant differences.

The annotated linear maps revealed that the mitochondrial genomes of SH, pm2, and XJ1 exhibited an inverted organization relative to other specimens (Fig. 2). Supplementary Material 4 contains the circular annotation maps. Plant pm1 has 68 annotated genes, including 39 CDS, 26 tRNAs, and three rRNAs. The remaining seven plants each have 67 annotated genes, including 39 CDS, 25 tRNAs, and three rRNAs, with the main difference

Table 2 Statistical information on the mitochondrial genome characteristics and sample sources of 12 plant specimens

ID	Paired-end reads	Length of mitochondrion (bp)	GC%	Accession number
sxupm2407041(pm1)	17,183,500	68695	42.1%	This study
sxupm2407042(pm2)	14,518,794	68719	41.8%	This study
sxuzm2406131(zm1)	13,144,812	69814	42.1%	This study
sxuzm2406132(zm2)	14,851,580	70447	42.2%	This study
sxuxm2407072(xm2)	17,212,080	67229	42.2%	This study
sxuxm2407073(xm3)	17,941,748	70015	42.2%	This study
sxuSH241014(SH)	17,969,404	67159	41.9%	This study
sxuXJ2410111(XJ1)	15,526,470	66695	41.8%	This study
ON406421	-	67738	42.0%	ON406421.1 (GenBank)
NC_005255	-	67737	42.0%	NC_005255.1 (GenBank)
NC_017598	-	80193	42.0%	NC_017598 (GenBank)
MW556320	-	60236	41.2%	MW556320 (GenBank)

being in the tRNA genes. Plant pm1 has 26 tRNA genes, one more tRNA-Arg than the others. The annotation results of pm1 are highly consistent with the reference genome. Additionally, among the 39 annotated protein-coding genes, *rps*3, *nd*4, *cox*1, *cox*2, *cob*, and *atp*9 genes contain intron regions. A structural difference was noted in the atp9 gene: the ortholog in plant pm1 possesses two introns, whereas in plant SH, it contains only one. Furthermore, homology comparison revealed that the *cox*1 gene in plant SH harbors seven introns. Validation using samtools shown an average read coverage of 66.219 for the SH sample, with *cox*1 gene region averaging 35, thereby meeting the quality standards for mitochondrial genome assembly [42, 43]. Analysis of rRNA genes revealed that the 5S Ribosomal RNA (*rrn*5) and small subunit ribosomal RNA genes (*rns*) genes were highly conserved in length across all specimens. In contrast, the large subunit ribosomal rRNA gene (*rnL*) exhibited length and positional variation and contained the highest number of introns. Comparative analysis with the reference genome revealed that the gene in this study lacks a segment of the intronic region. Through comparative homology analysis, the sample was found to be devoid of the *rnL*3, *rnL*4, and *rnL*9 gene segments annotated in the reference genome. A detailed presentation of the *rnL* gene alignment results is shown in Supplementary Material 5. In addition, the *rnL* gene in plant xm2 was found to contain an intergenic sequence.

Codon bias analysis

Cluster analysis of RSCU values for 12 mitochondrial genomes, as shown in Fig. 3, indicates significant results where AU p-values and BP values are greater than 95. From the perspective of codon bias characteristics, zm1 and pm1 plants show consistent codon preference with a significance of 100, while zm2 and xm2 plants exhibit highly similar codon preference with a significance of 98.

The RSCU value reflects the ratio of the actual usage frequency of each codon to its theoretical usage frequency. Codons with a ratio greater than one are high-frequency codons. The number of high-frequency codons possessed by the 8 plants ranges from 26 to 33, with zm1 plant having 33 high-frequency codons and pm2 plant having 26 high-frequency codons. The RSCU values for arginine AGA and proline CCA codons are the highest. The RSCU values for methionine AUG and tryptophan UGG codons are both one, indicating no preference. Among these eight plants, the amino acids with the highest content are arginine (Arg), leucine (Leu), and serine (Ser), each encoded by six codons. Next are alanine (Ala), glycine (Gly), proline (Pro), threonine (Thr), and valine (Val), each encoded by four codons. The lowest content is for methionine and tryptophan, each encoded by only one codon. Among the six codons encoding arginine, AGA has the highest RSCU value, indicating that AGA is the most commonly used codon in the mitochondrial genomes of these eight algae plants.

The eight samples exhibited CAI values of 0.169–0.177 and CBI values of (−0.065)-(−0.053), these results indicate no preferential use of optimal codons in the mitochondrial genomes. The GC3s values (x-axis) ranged from 38.5% to 40.10%, and the GC12 values (y-axis) ranged from 37.56% to 38.00%. The data fitting between GC12 and GC3s yielded the function $y = 0.0049x + 0.3759$, with a coefficient of determination (R^2) of 0.0003 and a *P*-value of 0.9670, demonstrating no significant linear correlation. In contrast, the Pearson correlation between ENC and GC3s was 0.9613 ($P = 0.00014$), confirming a significant positive relationship. As shown in Fig. 4a, all eight plants are distributed near the standard curve and the maximum vertical distance between the sample points and the standard curve is 0.095913, indicating that their codon bias is mainly influenced by changes in the genes themselves. Following the calculation, SH and pm1 shown the closest proximity to the curve, with vertical distances to the standard curve measuring 0.000571 for SH and 0.001845 for pm1. Additionally, the pm1 and zm1 plants almost overlap, which is consistent with the clustering results reflected in the clustering diagram.

Figure 4b shows the specific GG content, AT offset, and GC offset of 8 plants. It can be seen that the GC content is distributed between 41.8% and 42.2%, which is noticeably slightly lower than 50%. The GC offset is greater than the AT offset, with the AT offset ranging from −0.257 (xm2) to 0.257 (SH), and the GC offset ranging from −0.900 (pm2) to 0.900 (pm1).

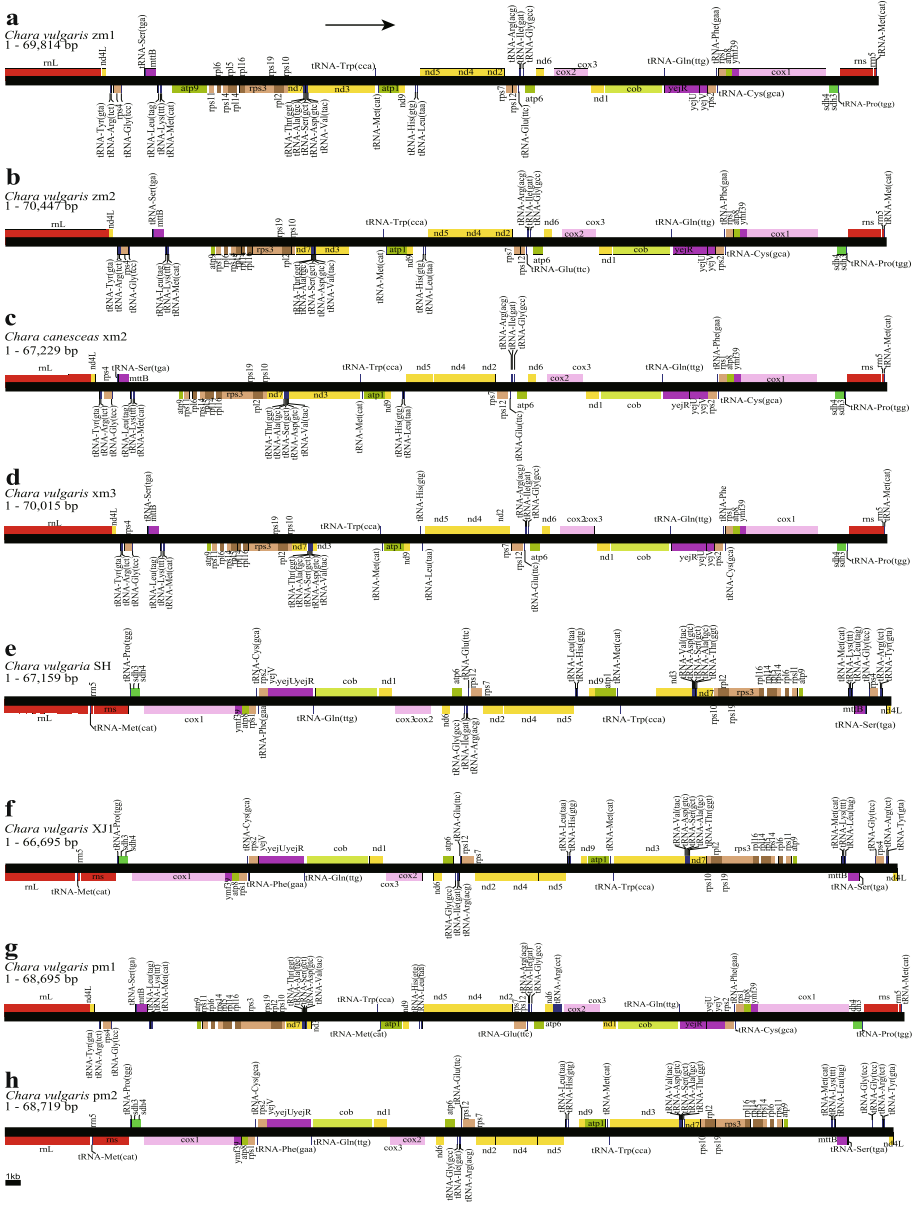

Fig. 2 Annotated linear map of mitochondrial genomes from 8 *Chara* samples. **a** zm1 sample. **b** zm2 sample. **c** xm2 sample. **d** xm3 sample. **e** SH sample. **f** XJ1 sample. **g** pm1 sample. **h** pm2 sample

Phylogenetic tree analysis

Phylogenetic relationships were inferred using the MrBayes method based on two distinct datasets: (i) mitochondrial amino acid sequences from 12 species, and (ii) a concatenated alignment of protein-coding and RNA sequences. Both analyses were rooted with two Coleochaetophyceae species (*C. orbicularis*, *C. nitellarum*), and the resulting topologies are presented in Fig. 5a and b, with all branches possessing support values > 50% displayed. The tree structures of the two are completely identical and have similar support rates. Phylogenetic analyses resolved the eight plant samples as a monophyletic group. (Fig. 5b). Within this clade, internal relationships were well-supported: SH and xm3 formed a strongly supported cluster (posterior probability, PP = 0.995), as did XJ1 and pm2 (posterior probability, PP = 0.65). In contrast, zm2, pm1, xm2, and zm1 were each resolved as distinct lineages, with PPs of 0.594, 1, 0.53, and 1, respectively. This topology was inferred from a concatenated mitochondrial amino acid dataset under a Bayesian framework, where the pm1 branch received a PP of 0.583.

The STAG analysis of mitochondrial homologous gene sequences generated a species phylogenetic tree (Fig. 5c), which displays branch support values greater than 50% and obtained topologically consistent evolutionary trees.

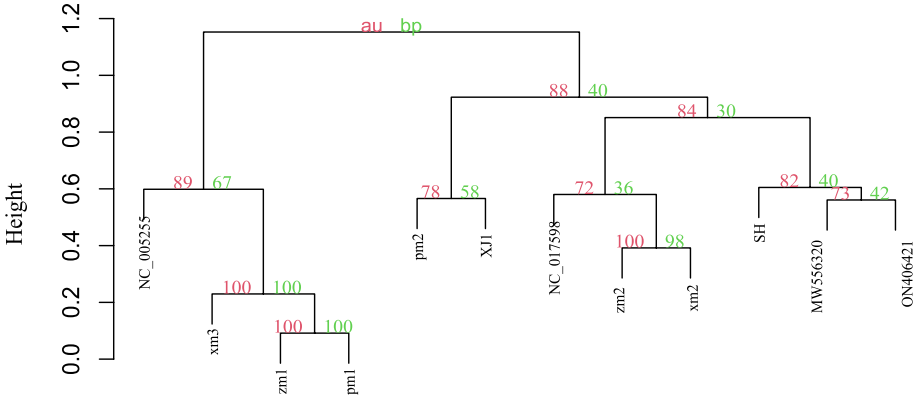

Fig. 3 RSCU-based cluster analysis with p – values (%)

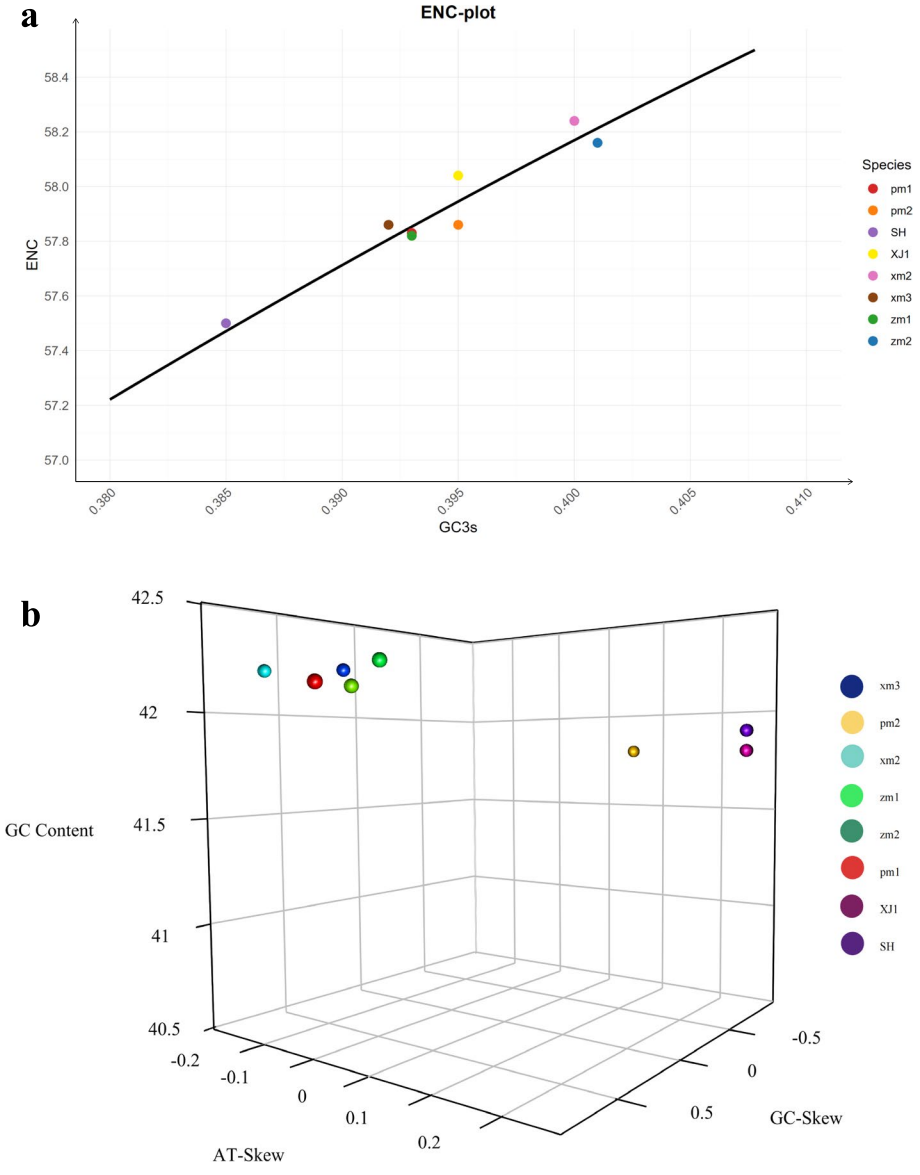

Fig. 4 ENC-plot and GC content of *Chara*. **a** Relation between GC3 and ENC (ENC were plotted against GC content at the third codon position and the expected ENC from GC3 are shown as a solid line). **b** Three-dimensional scatter plot of the AT-skew, GCskew, and GC content from 8 mitochondrion in *Chara*

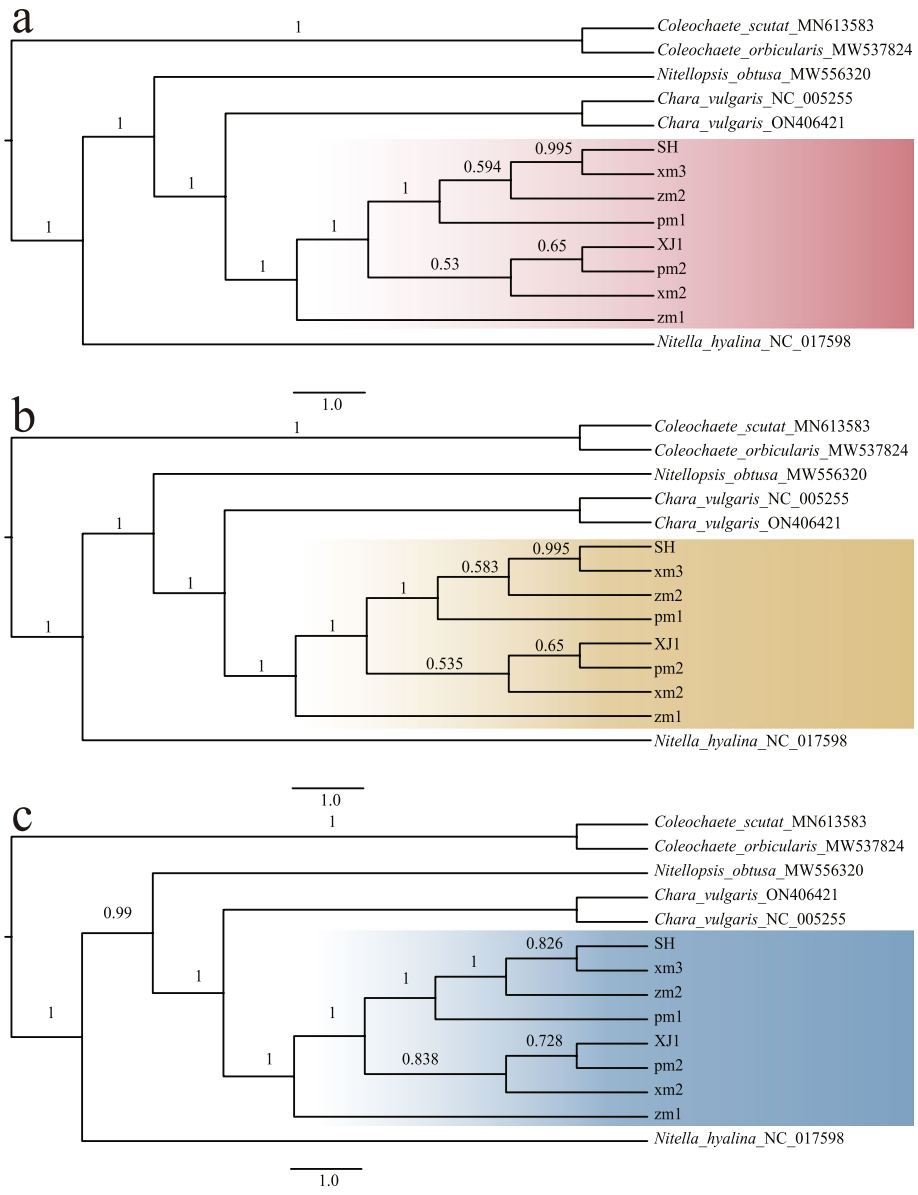

Fig. 5 Phylogenetic trees based on amino acid sequences, protein-coding sequences, and RNA concatenated sequences and species trees based on orthologous gene sequences (Support values for all analyses are noted as: Bayesian posterior probabilities). **a** MrBayes tree based PCGs-RNA sequences. **b** MrBayes tree based on amino acid sequences. **c** species tree based on orthologous gene sequences

Table 3 Phylogenetic tree evolutionary rate statistics

	Signal-to-noise	Treeness	Relative composition variability	Evolution rate
*rbc*L	3.678	0.617	0.168	0.020
PCGs-RNA	8.161	0.766	0.094	0.439
AA	8.151	0.765	0.094	0.439
STAG	19.723	0.880	0.045	0.033

The branch support rates shown by the species count differ slightly from those of the gene tree. For example, the posterior probability supporting the clustering of XJ1 plant and pm3 plant reaches 0.728, which is higher than the corresponding value in the gene tree. In contrast, the posterior probability supporting the clustering of SH plant and xm3 plant is 0.826, which is lower than the corresponding value in the gene tree. We performed evolutionary rate statistics on 4 different phylogenetic trees, and the results are shown in Table 3. The phylogenetic trees constructed based on amino acid sequences and protein-coding and RNA concatenated sequences have higher evolutionary rates as shown in Table 3. Among the phylogenetic trees reconstructed in this study, the one based on orthologous gene sequences exhibited the highest degree of resolution and topological stability. Consequently, it is considered to provide the most reliable

inference of evolutionary relationships. In contrast, the two gene trees demonstrated significantly higher evolutionary rates than the species tree.

Discussion

Morphological and molecular phylogenetic analysis

Our morphological and phylogenetic analysis shows that the eight strains in this study exhibit diverse morphologies. Among them, the xm2 strain shows significant morphological differences from the other plants and also has a distinct evolutionary position. The shape and number of spine cells, the morphology of stipulodes, and the structure of the reproductive organs of the xm2 strain are very special. The phylogenetic tree constructed based on *rbc*L gene indicates that the taxonomic position of the xm2 plant is relatively special. The results show that the xm2 plant is closely related to *C. altaica* and *C. canescens* with a high support rate of 0.997/84. Analysis of the *rbc*L sequences revealed that *C. altaica* and parthenogenetic specimens of *C. canescens* have identical *rbc*L sequences [44]. Hidetoshi Sakayama demonstrated that *C. altaica* and parthenogenetic populations of *C. canescens* were indistinguishable based on *rbc*L gene sequences [45]. Morphological evidence is congruent with the molecular data; specifically, xm2 resembles *C. canescens* in being dioecious, whereas *C. altaica* is monoecious. Therefore, based on the molecular and morphological evidence presented, xm2 is identified as *C. canescens*. In the phylogenetic tree constructed from mitochondrial genomes, the taxonomic placement of xm2 differs from the aforementioned conclusions. This discrepancy may be attributed to two factors. First, no mitochondrial genome data are available for *C. canescens*, preventing its inclusion in the phylogenetic reconstruction. Second, whether this inconsistency relates to the conserved nature of charophyte mitochondrial genomes requires further investigation with the inclusion of more diverse species. The morphologies of the other seven plants are generally similar, but there are certain differences in stipulodes size, bract length, and branch cortex. Among them, the pm1 and pm2 plants have the same habitat, and their morphologies have almost no change. However, the morphological differences between zm1 and zm2 plants, which also have the same habitat, are more significant. Based on the phylogenetic analysis, pm1, pm2, zm1, and zm2 were all identified as *C. vulgaris*. However, within this major clade, pm1 and pm2 did not cluster together, nor did zm1 and zm2, indicating a lack of strong positive correlation between morphological and molecular data for these four samples. Since the genotype responsible for bract length was not identified, we speculate that the morphological differences between zm1 and zm2 may be attributed to developmental plasticity, where local environmental factors during growth lead to distinct phenotypes. The longer bracts observed in zm2 suggest that this individual may have developed in a location with stronger water flow, potentially requiring enhanced physical protection for its antheridia and oogonia [46]. The phylogenetic tree constructed based on *rbc*L gene shows that the genus *Chara* is not supported as a monophyletic group, which is consistent with previous research results [47]. The pm1, pm2, zm1, zm2, SH, XJ1, and xm3 accessions form a well-supported clade with *C. vulgaris*. Within this clade, SH exhibits the greatest genetic distance from *C. vulgaris*, whereas pm2 and xm3 are the most closely related, albeit with weak statistical support.

There are obvious differences in the morphology of stinging cells, reproductive structures, and stipulodes morphology among the plants of the *Chara* genus. The above three morphological characteristics can be used as characteristic morphologies for species identification. In addition, bracts and the cortex of small branches can be used as supplementary morphologies. The genus *Chara* exhibits considerable interspecific morphological divergence. Notably, significant intraspecific variation is observed in some species and the morphology of *C. vulgaris* has been documented to vary considerably across different geographic origins and environmental gradients [5]. Despite occupying congruent evolutionary positions, their precise taxonomic delineation requires molecular data for resolution, with morphological characteristics serving an ancillary role in species identification.

Analysis of mitochondrial genome characteristics

In this study, the mitochondrial genome sizes of several species ranged from 66.7 kbp to 70.4 kbp. Comparative genomic analyses of published mitochondrial genomes across algal lineages reveal that the charophyte *Chara* possesses a genome of relatively compact size. The largest sequenced mitochondrial genome among algal species is that of *Chlorokybus atmophyticus* Geitler, at 201.8 kbp, while the smallest is *Mesostigma viride* Lauterborn, at 42.4 kbp [16]. This study annotated 39 CDS in the mitochondrial genome, which is 1–7 more than the average CDS number in the mitochondrial genomes of other algae including cryptophyte, diatoms, brown algae, and red algae [48]. This suggests that the mitochondrial genome of Charophytes contains more information and may regulate more complex life activities.

The functional gene sequences of plant mitochondrial genomes are highly conserved, but there are differences in the number, location, and order of functional genes among different species [49]. In this study, the 8 plants belong to *Chara*, and most of the annotated functional gene sequences are conserved. However, there are differences in the length and gene structure of genes encoding ATP synthesis and energy metabolism-related genes, and cytochrome c oxidase-related genes. We found that the

lengths of the *atp*9 and *cox*1 genes differ among different plants, and the number of introns contained in the genes also varies. Studies have shown that genes encoding components involved in cytochrome c biosynthesis in *Chara* mitochondrial DNA are very common in land plants but rarely found in algae [16]. Compared with higher plants, the mitochondrial genomes of Charophyta plants are more evolutionary conserved. Both hornworts and Charophyta mitochondrial genomes contain the *rrn*5 gene, but this gene is usually lost in vascular plant mitochondrial genomes [50]. However, the mitochondrial genomes of Charophyta plants are not as highly dynamic and open as those of hornworts. In many flowering plants, the *atp*9 gene in their own mitochondrial genomes has degenerated or been completely lost, while instead the nuclear genome contains intact, functional copies of the *atp*9 gene [51]. In addition, many *rpl* and *rps* genes (such as *rpl*2, *rpl*5, *rps*1, *rps*2, *rps*7, *rps*10, *rps*11, *rps*12, *rps*14, etc.) have been lost or transferred to the nucleus in the common ancestor of angiosperms or in different lineages after that [52].

We further analyzed the number and distribution of introns in the mitochondrial genome and found that the distribution of introns in the mitochondrial genome also varies among the 8 plants. In this study, the number of intron-containing genes in the 8 plants ranges from 5 to 7. Such differences also exist between bryophytes and vascular plants. Overall, the number of intron-containing genes in algae is generally less than that in ferns and terrestrial plants [53]. In bryophytes, the intron content varies greatly, ranging from 1 to 9, while in vascular plants, the intron content ranges from 0 to 5 [54]. However, there is DNA sequence transfer between nuclear genomes and mitochondrial genomes in ferns, terrestrial plants, and even green algae, such as the phenomenon of chloroplast DNA being transferred to mitochondrial genomes [55, 56]. No DNA transfer between the nucleus and cytoplasm occurred in the 8 plants in this study.

From the perspective of GC content, the GC content of all 8 plants was slightly lower than 50%, which is consistent with the relatively low GC content in plant mitochondrial genomes [54]. This indicates that the mitochondrial genomes of Characeae plants tend to be rich in A/T nucleotides, but this tendency is not significant, and the AT content is not as high as 70.4% in red algal mitochondrial genomes [57]. In line with prior research, all investigated species of the Charophyta demonstrated an absence of GC-biased codon usage [58]. And there is no significant difference in GC content between different strains, so the conservatism of mitochondrial genomes can also be seen from the GC content.

The results of codon usage bias analysis reflect both the consistency of the mitochondrial genome and certain differences. The results show the ratio of the actual usage frequency of each codon to its theoretical usage frequency. When RSCU is greater than 1, it indicates that the usage frequency is relatively high, and such codons are called high-frequency codons. When RSCU equals 1, the codon has no preference [52]. According to the clustering results of relative synonymous codon usage, the relative synonymous codon usage of the three plants xm3, zm1, and pm1 is small, with values of 30, 33, and 31 respectively. The relative synonymous codon usage of zm2 and xm2 plants also differs slightly, with the number of high-frequency codons being 28 and 29 respectively. Our results confirm that T- and A-ending codons serve as the preferred codons in the mitochondrial genomes of the eight plants examined, which was similar with previous studies [53, 54]. A specific analysis of the usage preference of each amino acid for different codons in the mitochondrial genomes of the 8 plants shows consistency, among which the usage of arginine, tryptophan, and leucine is the highest. The AGA codon encoding arginine is the most commonly used codon in the mitochondria of these 8 plants.

To further illustrate the usage preference of synonymous codons, an ENC-plot was drawn based on the GC content at the third position of synonymous codons (GC3s). The preference of plant mitochondrial genome codons may be driven primarily by mutation or natural selection [59]. Previous studies have shown that in Chlorophyta, 11 out of 12 species had their mitochondrial genes distributed well below the expected curve. On the contrary, in Streptophyte, all 24 species in study had their mitochondrial genes distributed around the expected curve, one of the species is *C. vulgaris* [58]. It can be seen that different plants experience inconsistent evolutionary driving forces during evolution. The ENC-plot in this study showed that all eight sample points were distributed near the expected curve, with relatively small vertical distances, suggesting that codon usage bias in the mitochondria is substantially influenced by mutation. However, the quantitative finding of no correlation between GC12 and GC3s indicates that codon usage preference is hardly driven by mutation pressure and is instead shaped primarily by natural selection. The calculation of the ENC value depends on both GC12 and GC3s. This study, GC12 remained nearly constant and did not vary with GC3s. Consequently, the observed ENC values closely matched the expected values predicted solely by GC3s, creating an "illusion of neutrality". To investigate whether selection for translational optimization exists, we analyzed CAI and CBI values. The results indicate the absence of such positive natural selection in this group. The codon usage pattern is thus neither dominated by strong mutation pressure nor influenced by selection for translational efficiency. We therefore speculate that intergenic variation in codon usage

primarily arises from genetic drift or other weak selective pressures, representing a nearly neutral state. To confirm this codon usage pattern, we further calculated the correlation coefficient between ENC and GC3s, obtaining a significant positive correlation. This result supports the inference of a near-neutral mode. In the absence of sustained mutational or translational optimization pressure, differences in codon usage among genes (particularly in GC3s) are governed by random genetic drift. This leads to the strong statistical correlation between ENC and GC3s and causes the observed values to cluster near the neutral expectation curve.

However, there is little discussion on the mitochondrial genomes of Charophyta plants at present. More extensive patterns of codon preference should be sought to provide more comprehensive and in-depth insights into revealing the phylogenetic relationships of plant groups and understanding their evolutionary history.

Comparative analysis of phylogenetic trees constructed based on different datasets

Phylogenetic trees were constructed using three different datasets from mitochondrial genomes: amino acid sequence datasets, concatenated datasets of protein-coding genes and RNA sequences, and homologous gene datasets. *Coleochaete orbicularis* and *C. nitellarum*, two species of Coleochaetophyceae, were selected as outgroups for all three datasets. The first two datasets were analyzed using bayesian inference, while the latter was analyzed using the STAG method. The STAG method does not consider the copy number of genes in species, and on actual species datasets, its results are superior to other methods used for inferring species numbers [40]. Overall, the evolutionary rate of gene trees is higher than that of species trees. Despite utilizing distinct datasets, the three phylogenetic analyses yielded highly congruent topologies. As exemplified by the identical tree structures in Figs. 5a and b, all analyses consistently and unequivocally supported the eight plant accessions as a monophyletic group, with 100% bootstrap support. This result reflects that the eight samples share a similar evolutionary trajectory, with xm2 also exhibiting an evolutionary path similar to that of *C. vulgaris*. In this study, the phylogenetic trees reconstructed from different mitochondrial genomic datasets of the eight samples demonstrated high topological congruence. This result indicates that the mitochondrial genomes of the eight plants contain consistent phylogenetic signals, showing similar evolutionary relationship between the eight plants in the study and *C. vulgaris*. However, certain branch nodes received only moderate statistical support. We discuss the potential influences of methodological choices and dataset composition, proposing that the high sequence conservation of mitochondrial genomes and incomplete taxon sampling may contribute to the weak phylogenetic signal at these nodes. Consequently, expanded taxonomic sampling and multi-model phylogenetic comparison should be prioritized in future studies [60, 61].

The congruence of the phylogenetic tree indicates that Characeae plants have consistent phylogenetic signals during evolution, and the mitochondrial genome as a whole record this evolutionary relationship, with different genes within it reflecting the same phylogenetic relationship, further illustrating the conservatism of mitochondrial genome evolution. This phylogenetic framework lays a reliable foundation for subsequent studies (such as evolutionary rate analysis, ancestral genome reconstruction, etc.) [62]. Although the consistency of the genomic results supports analyzing the mitochondrial genome as a whole, it has certain limitations. The phylogenetic trees exhibited limited nodal support at certain branches. Specifically, the xm2 and zm2 clades were not strongly supported (bootstrap value = 0.5) and exhibited discordance among the three inferred trees. The genomic phylogenetic analysis in this study only includes three genera in Characeae and has not covered all major clades of *Chara*, and there are only two high-quality mitochondrial genomes of *Chara* genus available: *C. braunii* and *C. vulgaris*. This has led to certain limitations in reconstructing phylogenetic trees, making it impossible to infer more reliable evolutionary relationships through more datasets. Compared with the phylogenetic tree constructed based on a single gene (*rbc*L), the taxonomic position of the xm2 plant is significantly different, which is because the xm2 plant lacks mitochondrial genome data from closely related species. Extensive simulation and empirical studies have shown that sparse taxon data will reduce the resolution and support of phylogenetic relationships and increase the risk of erroneous inferences caused by long-branch attraction [63, 64]. Although the structure of the genomic phylogenetic tree in this study is supported under the current data, future research urgently needs to include more key clades to verify and improve the current phylogenetic framework [65].

Conclusion

The mitochondrial genome architectures of eight Characeae species were investigated and the phylogenetic relationships were evaluated based on various datasets. The assembled mitochondrial genomes of the eight examined species range from 66.7 to 70.4 kbp, with size variations primarily attributed to intergenic spacer regions. Our analyses revealed that Charophyte mitochondrial genomes, representing a key transitional stage from aquatic to terrestrial habitats, lack a distinct AT bias. Instead, their GC content is notably higher and appears to be undergoing a gradual accumulation, a trend more aligned with higher plants. Furthermore,

while Charophyte mitochondrial genomes are more complex than those of other algae, they exhibit greater evolutionary conservation compared to their counterparts in higher plants. This study first analyzes the codon usage bias characteristics of Charophyte mitochondrial genomes and shows that the codon usage bias of the eight plants is consistent, among which the AGA codon encoding arginine is the most commonly used. Integrated analysis of the codon adaptation index, GC12-GC3s data fitting, and the effective number of codons collectively demonstrates that codon usage in the eight mitochondrial genomes studied is not significantly influenced by translational optimization selection or directional mutation pressure. Instead, the observed preference pattern is more consistent with the characteristics of nearly neutral evolution, which is primarily driven by genetic drift. The topology of the phylogenetic tree is closely related to the number of taxonomic groups and the number of genes included in the dataset. The single-gene (*rbc*L) dataset provides sufficient sample size, enriches the species information within the genus *Chara*, and clarifies the taxonomic status of the xm2 plant. However, the mitochondrial genome dataset fails to cover enough Charophyte taxonomic groups, resulting in an inability to reflect the true closely related taxa of xm2. Phylogenetic relationships reflected by different mitochondrial genome datasets are highly consistent, indicating that the mitochondrial genomes of Characeae plants contain consistent phylogenetic signals, revealing reliable and similar evolutionary relationships between the eight species in the study and *C. vulgaris*. In addition, the genomic data of this study greatly enriches the gene database of Charophyte plants, laying a foundational framework for expanding the dataset and further clarifying phylogenetic relationships in subsequent studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12788-7.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Authors' contributions

Y.H. Xie wrote and edited the manuscript text, completed data analysis and graph plotting and interpreted all the data and findings. H.L. Bai reviewed the manuscript. X.D. Liu reviewed the manuscript. J. Feng reviewed the manuscript. F.R. Nan reviewed and edited the manuscript and prepared investigation. S.L. Xie reviewed the manuscript and prepared resources, project administration, investigation, funding acquisition and conceptualization. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 32170204 to S.L. Xie, No. 32270220 and U22A20445 to J. Feng), the Water Conservancy Science and Technology Research and Promotion Projects in Shanxi province (2025GM38) to F.R. Nan, the Program of Sanjin Talent of Shanxi Province (No. SJYC2020502) to J. Feng, the Grant from Yuncheng Salt Lake Protection and Utilization Research Institute of Shanxi (No. YHYJ-2023003) to S.L. Xie, the Shanxi Provincial Key Research and Development Project (No. 202202140601019) to X.D. Liu.

Data availability

The raw data have been submitted to NCBI under accession number PX641307、PX593782、PX778740、PX778741、PX778742、PX778743、PX778744 and PX393130.

Declarations

Ethics approval, and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]School of Life Science, Shanxi University, Taiyuan 030006, China
[2]Shanxi Key Laboratory for Research and Development of Regional Plants, Shanxi University, Taiyuan 030006, China

Received: 18 December 2025 / Accepted: 20 March 2026
Published online: 27 March 2026

References

1. McCourt RM, Delwiche CF, Karol KG. Charophyte algae and land plant origins. Trends Ecol Evol. 2004;19:661–6.
2. Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, et al. The *Chara* genome: secondary complexity and implications for plant terrestrialization. Cell. 2018;174:448–64.
3. Lambert S. Stoneworts: their habitats, ecological requirements and conservation. Science Report No. SC030202. Bristol: Environment Agency; 2009. p. 15.
4. Bellino A, Baldantoni D. Biodiversity, ecology and distribution of Mediterranean charophytes in Southern Italy. Plants. 2023;12:3434.
5. Denys L, Christia C, Papastergiadou E. *Chara vulgaris*. In: Schubert H, et al. ed. Charophytes of Europe. Cham: Springer; 2024. p. 695–725.
6. Domozych D, Sørensen I, Popper ZA. Editorial: charophytes: evolutionary ancestors of plants and emerging models for plant research. Front Plant Sci. 2017;8:338.
7. Wood RD, Imahori K. A revision of the Characeae. Taxon. 1965;14:240–4.
8. Domozych DS, Bagdan K. The call biology of charophytes: exploring the past and models for the future. Plant Physiol. 2022;190:1588–608.
9. Kranz HD, Mikš D, Siegler ML, Capesius I, Sensen CW, Huss VAR. The origin of land plants: phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. J Mol Evol. 1995;41:74–84.
10. Bhattacharya D, Medlin L. Algal phylogeny and the origin of land plants. Plant Physiol. 1998;116:9–15.
11. Marin B, Melkonian M. Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. Protist. 1999;150:399–417.
12. Delwiche CF, Karol KG, Cimino MT, Sytsma KJ. Phylogeny of the genus *Coleochaete* (Coleochaetales, Charophyta) and related taxa inferred by analysis of the chloroplast gene rbcL. J Phycol. 2002;38:394–403.
13. Karol KG, McCourt RM, Cimino MT, Delwiche CF. The closest living relatives of land plants. Science. 2001;294:2351–3.
14. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot. 2012;99:349–64.

15. Chapman RL, Buchheim MA, Delwiche CF, Friedl T, Huss VAR, Karol KG, et al. Molecular Systematics of the Green Algae. In: Soltis DE, Soltis PS, Doyle JJ, editors. Molecular Systematics of Plants II: DNA Sequencing. Boston, MA: Springer, US; 1998. p. 508–40.
16. Burger G, Nedelcu AM. Mitochondrial Genomes of Algae. In: Bock R, Knoop V, editors. Genomics of Chloroplasts and Mitochondria. Dordrecht: Springer, Netherlands; 2012. p. 127–57.
17. Oudot-Le Secq MP, Fontaine JM, Rousvoal S, Kloareg B, Loiseaux-de Goër S. The complete sequence of a brown algal mitochondrial genome, the Ectocarpale *Pylaiella littoralis* (L.) Kjellm. J Mol Evol. 2001;53:80–8.
18. Turmel M, Otis C, Lemieux C. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc Natl Acad Sci U S A. 2002;99:11275–80.
19. Mazumder GA, Uddin A, Chakraborty S. Analysis of codon usage bias in mitochondrial CO gene among platyhelminthes. Mol Biochem Parasitol. 2021;245:111410.
20. Zhao D, Ye T, Gao F, Jakovlić I, La Q, Tong Y, et al. PhyloSuite v2: The development of an all-in-one, efficient and visualization-oriented suite for molecular dating analysis and other advanced features. iMeta. 2025;4:e70095.
21. Turmel M, Otis C, Lemieux C. The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell. 2003;15:1888–903.
22. Song Q, Liu X, Nan F, Liu Q, Lv J, Feng J, et al. *Chara zhengzhouensis* (Characeae, Charophyta), a new freshwater algal species described from North China. Diversity. 2022;14:864.
23. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
25. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualisation of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.
26. Cahoon AB, Nauss JA, Stanley CD, Qureshi A. Deep transcriptome sequencing of two green algae, *Chara vulgaris* and *Chlamydomonas reinhardtii*, provides no evidence of organellar RNA editing. Genes. 2017;8:80.
27. Rose R, Golosova O, Sukhomlinov D, Tiunov A, Prosperi M. Flexible design of multiple metagenomics classification pipelines with UGENE. Bioinformatics. 2019;35:1963–5.
28. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–6.
29. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.
30. Lehwark P, Greiner S. GB2sequin - a file converter preparing custom GenBank files for database submission. Genomics. 2019;111:759–61.
31. Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, et al. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. Nat Plants. 2020;6:95–106.
32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
33. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol. 2018;35:2582–4.
34. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56:564–77.
35. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
36. Lanfear R, Frandsen P, Wright A, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol. 2017;34:772–3.
37. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
38. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
39. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
40. Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. bioRxiv. 2018;267914. https://doi.org/10.1101/267914.
41. Wright F. The 'effective number of codons' used in a gene. Gene. 1990;87:23–9.
42. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10:giab008.
43. Mower JP, Case AL, Floro ER, Willis JH. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol. 2012;4:670–86.
44. Kato S, Sakayama H, Morishima H, Sano S, Oomori Y, Kato N, et al. Morphology and molecular phylogeny of *Chara altaica* (Charales, Charophyceae), a monoecious species of the section *Desvauxia*. Cytologia. 2010;75:211–20.
45. Sakayama H, Omori Y, Nozaki H, Kato S, Kamiya K, Watanabe MM. Taxonomy, new distribution, and DNA barcoding of the rare alga *Chara altaica* (Charales, Charophyceae) in Japan. J Asia-Pac Biodivers. 2024;17:712–21.
46. Li P, An Z, Xu N, Li J, Li Q, He C. Phenotypic plasticity and stability in plants: genetic mechanisms, environmental adaptation, evolutionary implications, and future directions. Plant Cell Environ. 2025;48:5847–60.
47. Meiers S, Rootes WL, Proctor VW, Chapman RL. Phylogeny of the Characeae (charophyta) inferred from organismal and molecular characters. Arch Protistenkd. 1997;148:308–17.
48. Kim JI, Yoon HS, Yi G, Shin W, Archibald JM. Comparative mitochondrial genomics of cryptophyte algae: gene shuffling and dynamic mobile genetic elements. BMC Genomics. 2018;19:275.
49. Palmer JD, Herbon LA. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28:87–97.
50. Xue J-Y, Liu Y, Li L, Wang B, Qiu Y-L. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. Curr Genet. 2010;56:53–61.
51. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424:197–201.
52. Adams KL, Daley DO, Qiu Y-L, Whelan J, Palmer JD. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature. 2000;408:354–7.
53. Xu W, Xing T, Zhao M, Yin X, Xia G, Wang M. Synonymous codon usage bias in plant mitochondrial genes is associated with intron number and mirrors species evolution. PLoS ONE. 2015;10:e0131508.
54. Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. Mol Biol Rep. 2009;36:2039–46.
55. Hao W, Palmer JD. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. Proc Natl Acad Sci U S A. 2009;106:16728–33.
56. Tsunewaki K. Interorganellar DNA transfer in wheat: dynamics and phylogenetic origin. Proc Jap Acad Ser B - Phys Biol Sci. 2011;87:529–49.
57. Patil MP, Oktavitri NI, Kim YR, Yoon S, Lee IC, Kim JO, et al. Complete mitochondrial genome and its phylogenetic position in red algae *Fushitsunagia catenata* from South Korea. Life. 2024;14:534.
58. Wang B, Yuan J, Liu J, Jin L, Chen J-Q. Codon usage bias and determining forces in green plant mitochondrial genomes. J Integr Plant Biol. 2011;53:324–34.
59. Salim H, Cavalcanti A. Factors influencing codon usage bias in genomes. J Braz Chem Soc. 2008;19:257–62.
60. Xue T-T, Janssens SB, Liu B-B, Yu S-X. Phylogenomic conflict analyses of the plastid and mitochondrial genomes via deep genome skimming highlight their independent evolutionary histories: A case study in the cinquefoil genus *Potentilla* sensu lato (Potentilleae, Rosaceae). Mol Phylogenet Evol. 2024;190:107956.
61. Feist M, Grambast-Fessard N. The Genus Concept in Charophyta: Evidence from Palaeozoic to Recent. In: Riding R, editor. Calcareous Algae and Stromatolites. Berlin: Springer; 1991. p. 189–203.
62. Jeffroy O, Brinkmann H, Delsuc F, Philippe H. Phylogenomics: the beginning of incongruence? Trends Genet. 2006;22:225–31.
63. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. Syst Biol. 1978;27:401–10.

64. Graybeal A. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst Biol. 1998;47:9–17.
65. Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. Increased taxon sampling is advantageous for phylogenetic inference. Syst Biol. 2002;51:664–71.