RESEARCH

De novo assembly of the complete mitochondrial genome of sweet potato (*Ipomoea batatas* [L.] Lam) revealed the existence of homologous conformations generated by the repeat-mediated recombination

Zhijian Yang[1,2†], Yang Ni[2*†], Zebin Lin[1], Liubin Yang[1], Guotai Chen[1], Nuerla Nijiati[1], Yunzhuo Hu[1] and Xuanyang Chen[1,2,3*]

Abstract

Sweet potato (*Ipomoea batatas* [L.] Lam) is an important food crop, an excellent fodder crop, and a new type of industrial raw material crop. The lack of genomic resources could affect the process of industrialization of sweet potato. Few detailed reports have been completed on the mitochondrial genome of sweet potato. In this research, we sequenced and assembled the mitochondrial genome of sweet potato and investigated its substructure. The mitochondrial genome of sweet potato is 270,304 bp with 23 unique core genes and 12 variable genes. We detected 279 pairs of repeat sequences and found that three pairs of direct repeats could mediate the homologous recombination into four independent circular molecules. We identified 70 SSRs in the whole mitochondrial genome of sweet potato. The longest dispersed repeat in mitochondrial genome was a palindromic repeat with a length of 915 bp. The homologous fragments between the chloroplast and mitochondrial genome account for 7.35% of the mitochondrial genome. We also predicted 597 RNA editing sites and found that the *rps*3 gene was edited 54 times, which occurred most frequently. This study further demonstrates the existence of multiple conformations in sweet potato mitochondrial genomes and provides a theoretical basis for the evolution of higher plants and cytoplasmic male sterility breeding.

Keywords: *Ipomoea batatas*, Mitochondrial genome, De novo assembly, Repeat-mediated recombination, RNA editing events

†Zhijian Yang and Yang Ni contributed equally to this work.

*Correspondence: ny_work@126.com; cxy@fafu.edu.cn

[2] College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China
[3] Fujian Provincial Key Laboratory of Crop Breeding by Design, Fuzhou, Fujian, China
Full list of author information is available at the end of the article

Introduction

Sweet potato [(*Ipomoea batatas* (L.) Lam.], which belongs to the botanical family Convolvulaceae [1], is an important food crop, an excellent fodder crop, and a new type of industrial raw material crop [2, 3]. The Food and Agriculture Organization of the United Nations (FAO) considers sweet potato one of the important crops in solving food shortage and energy problems in

Yang *et al. BMC Plant Biology*   (2022) 22:285

Page 2 of 12

the twenty-first century. In 2020, 7.40 million hectares of sweet potato were planted worldwide, with a total production of 89.49 million tons (https://www.fao.org/faostat/zh/#data,2022). The production and storage process of sweet potato is green and pollution-free, which is why sweet potato is considered "the most ideal food in the 21st century" [3]. In addition to being rich in starch, sweet potatoes also contain protein, minerals, and trace elements, especially a large amount of carotenoids and anthocyanins [4, 5]. These compounds have important health functions for human health, and recently, many new varieties of sweet potato with high pigment content have been successfully bred and applied in production [6, 7]. As sweet potatoes can be harvested for fresh consumption or for processing various food products, its preservation and processing technologies have been widely researched to improve the technology and efficiency of the sweet potato industry [8]. Sweet potato production is severely affected by biotic adversities, such as small weevil (weevil), SPVD virus disease, and root rot, as well as abiotic adversities such as drought and salinity [9–11], which can lead to yield losses and reduced quality of sweet potato. In the process of selecting and breeding new varieties with high resistance, many genes for good traits, especially genes for specific traits derived from wild species, are difficult to introduce into new varieties due to the complex genome [12]. With the development of bioinformatics, sweet potato has made great progress in genome, transcriptome, and small RNA analysis, laying the foundation for improving sweet potato traits and ensuring production safety [13–15].

Mitochondria are an essential organelle in eukaryotic cells and are a powerful tool for studying the origin of species, genetic diversity, and phylogenetics [4]. Mitochondrion and chloroplast are organelles with a semi-autonomous genetic system in higher plant cells, and they carry relevant genetic information. To date, 7367 chloroplast genomes and 1118 plastomes have been published in the NCBI database. However, only 441 plant mitochondrial genomes exist, according to the NCBI database (2022, April 23th, https://www.ncbi.nlm.nih.gov/genome/browse/#!/organelles/). Mitochondria play an important role in plant phylogeny and are widely used in phylogenetic and interspecies discrimination studies because their genetic system is relatively independent of the nucleus and relatively conserved [16]. In addition, studies have shown that mitochondria are closely related to cytoplasmic male sterility (CMS) and that self-incompatibility and hybrid incompatibility exist in sweet potato. Therefore, the study of the sweet potato mitochondrial genome can help identify and utilize wild seed resources of sweet potato, study the molecular mechanism of hybrid sterility, and provide services for sweet

potato variety improvement [17, 18]. The study of these genomes has lagged behind that of chloroplast and plastid genomes because of the complex structure of plant mitochondrial genomes. Currently, 15 mitochondrial genomes have been published in the convolvulaceae family, while the mitochondrial genomes of sweet potato remain unreported.

In this research, we sequenced and assembled the mitochondrial genome of the sweet potato and further investigated its substructure. We explored the existence of its mitochondrial genomic substructure through nanopore reads and PCR experiments. To better study its mitochondrial and chloroplast genome homologous fragments, we assembled its chloroplast genome by using the same set of data and completed a sequence similarity analysis. We also analyzed its mitochondrial genome for RNA editing events. This study further demonstrates the existence of multiple conformations in plant mitochondrial genomes and provides a theoretical basis for the evolution of higher plants and CMS breeding.

Material and methods
Plant materials, DNA extraction, and sequencing
Fresh leaves of sweet potato plants (JinShan 57) were collected in Fujian Agriculture and Forestry University (Longitude: 119.306239, Latitude: 26.075302). The leaves were cleaned with DEPC water and stored in a freezer at −80 °C. The DNA of sweet potato was extracted by using a DNA plant extraction kit (Tiangen, China). The short-paired reads were sequenced by Illumina HiSeq X ten (Illumina, Inc.; San Diego, CA, USA). For Oxford Nanopore sequencing, gTube was used to break genomic DNA into about 10 kb on average. The DNA library was taken, mixed with relevant reagents on board, and added to Flowcell for real-time single-molecule sequencing on the PromethION sequencer to obtain raw sequencing data.

Genome assembly, polish, and annotation
For the chloroplast genome assembly, we used GetOrganelle [19] to assemble the whole genome sequencing (WGS) data with default parameters directly. Then, the chloroplast genome fragments generated from GetOrganelle were used to filter the raw nanopore and illumina data. For the mitochondrial genome assembly, we first used the Nextdenovo (https://github.com/Nextomics/NextDenovo) software de novo assembly with the nanopore reads and selected them with the reference genome *Ipomoea nil* (NC_031158.1). The mitochondrial genome results were manually checked by drawing the dotplot by Gepard [20]. The illumina data were assembled into a uniting graph by GetOrganelle. The graph-based mitochondrial genome was visualized by Bandage and used to remove the chloroplast- and nuclear-derived uniting

Yang *et al. BMC Plant Biology* (2022) 22:285

Page 3 of 12

nodes manually. The repeat regions were solved with the long-read results from NextDenovo.

The chloroplast genome was annotated with the CPGAVAS2 [21] and checked by CPGview-RSG (http://www.herbalgenomics.org/cpgview/,unpublished). The PCGs and rRNA of mitogenome were annotated with Geseq [22] and BLASTN. The tRNAs were identified with tRNAscan [23] (version 1.4). The genome circle map of mitochondrial genome was visualized with OGdraw [24]. All the annotations of organelle genomes were reviewed carefully and manually corrected with Apollo software [25, 26].

Repeat and homologous DNA analysis
Microsatellite repeats were identified by MISA [27] with the parameters "1-10 2-6 3-5 4-5 5-5 6-5." Dispersed repeats were found with REPuter web server (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) with default parameters. Homologous DNA fragments were discovered between chloroplast genome and mitochondrial genome by BLASTN with the e-value of 1e-6 and word size 7 [28].

Repeat-mediated homologous recombination prediction and PCR amplification validation
BLASTN was used to detect the paired repeat sequences. The potential recombination was identified from repeat sequences. Each pair of repeats and its neighboring 1000 bp were extracted as two template sequences. After that, we formed the other two conformations after recombination by exchanging the neighboring 1000 bp sequences. The long reads were mapped to the template sequences and carefully checked to see which one can pass the repeats for four template sequences. The primers at the ends of repeats were designed with IDT web server and shown in Table S1. The PCR amplification protocol was 50 μl in total, consisting of 2 μl Template DNA, 2.5 μl forward primer, 2.5 μl reverse primer, 25 μl 2x Unique™ Taq Super Master Mix, and 18 μl ddH$_2$O. After initial denaturation at 94 °C for 3 min, PCR reactions were conducted for 35 cycles. Each cycle included denaturation at 94 °C for 10 s, annealing at 55 °C for 20 s, and elongation at 72 °C for 15 s. After the cycles ended, they were eventually extended for an additional 5 minutes.

Phylogenetic analysis and RNA editing sites prediction
Sixteen mitochondrial genomes from convolvulaceae species were chosen for phylogenetic analysis, and *Asclepias syriaca* (Mitochondrion: NC_022796.1) was set as the outgroup. The common genes of mitochondrial genomes were filtered with BLASTN and extracted, concatenated with Phylosuite [29]. Multiple sequence alignment was conducted by using MAFFT [30]. The

alignment results were calculated with MRBAYES [31]. Finally, the maximum-likelihood tree was visualized using iTOL (https://itol.embl.de/) [32]. We first downloaded the RNA-seq data from NCBI SRA database with the accession number DRR299538. We then decompressed this data using the SRAtoolkit [33] software to obtain the paired-end raw data. The protein-coding gene (PCG) sequences were extracted by using the Phylosuite software [29]. The RNA-seq data were mapped to all the PCGs of mitochondrial genome by using the BWA software [34]. The possible RNA editing sites were identified by using the Samtools program [35] according to the BAM file, and the locations with a coverage depth of more than 10x were selected.

Results
Mitochondrial genome assembly and annotation of sweet potato
In total, 10 GB Illumina reads and 10 GB nanopore were generated for genome assembly (SRA number: SRR17210632, SRR17210633). The plant mitochondrial genome exhibits a complex conformation due to a large number of repeat sequences with NGS data (Figure S1) [36]. The nanopore raw reads were assembled into a circular molecular structure. The single circle was used to resolve the repeats on the graph. We adopted a hybrid assembly model and presented its mitochondrial genome temporarily in the form of a molecular circle with 270,304 bp (Fig. 1). In total, we filtered 2.8 Gb nanopore reads and 278 Mb Illumina reads for organelle. The coverage depth of mitochondrial genome is 312.97 x for long reads and 332.9 x for short reads. The GC contents of the mitochondrial genome were 44.08%.

We annotated the mitochondrial genome, and the categorization of genes is shown in Table 1. The sweet potato mitochondrial genome has 23 unique core genes and 12 variable genes. The core genes consist of five ATP synthase genes (*atp*1, *atp*4, *atp*6, *atp*8, and *atp*9), nine NADH dehydrogenase genes (*nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *nad*7, and *nad*9), three cytochrome C biogenesis genes (*ccm*B, *ccm*C, *ccm*Fn), three cytochrome C oxidase genes (*cox*1, *cox*2, and *cox*3), ubiquinol cytochrome c reductase (*co*b), a transport membrane protein (*mtt*B), and a maturase (*mat*R). The variable genes consist of three large subunits of ribosome proteins (*rpl*5, *rpl*10, and *rpl*16), eight small subunits of ribosome proteins (*rps*1, *rps*3, *rps*4, *rps*10, *rps*12, *rps*13, *rps*14, and *rps*19), and one respiratory gene (*sdh*4). *ccm*Fc lost some part of exons and became a pseudogene. In all, 6 rRNAs and 16 tRNA were annotated in the sweet potato mitochondrial genome (Table 1). Among the core genes, the *ccm*Fc gene has become a pseudogene due to the loss of

Fig. 1 Representative genome map representing the mitochondrial genome circular molecule of *Ipomoea batatas*. The colored squares distributed inside and outside the circle represent different mitochondrial genes. Gene taxa of the same function are represented using the same color. In this case, the trans-shear genes nad1, nad2, nad5, etc. are represented as multiple exons

part of the first exon. No PCGs are in duplicated regions, and all PCGs are single-copy.

Repeats mediate the homologous recombination

A single circular molecule is not sufficient to display the plant mitochondrial genome. Repeated sequences may be able to mediate varying degrees of rearrangement of the genome. To further investigate the possible occurrence of homologous recombination in the mitochondrial genome of sweet potato, we detected 279 pairs of repeat sequences in the sweet potato mitochondrial genome by using BLASTN with the e-value 1*e*-5. On the basis of the long reads, each pair of repeat sequences was carefully checked, and we found that three pairs of reads may support homologous recombination (Table 2). The length of those repeats ranges from 62 to 253 bp. All the pairs of repeats were direct repeats, and each set of direct repeats may cause the mitochondria genome to form into two separate circular molecules according to our preliminary judgment.

To further verify these potential homologous recombinations, we designed primers at each end of the

Yang *et al. BMC Plant Biology* (2022) 22:285

Page 5 of 12

Table 1 Gene composition in the mitogenome of sweet potato

Group of genes		Name of genes
Core genes	ATP synthase	*atp1, atp4, atp6, atp8, atp9*
	Cytochrome c biogenesis	*ccm*B, *ccm*C, *ccm*Fn, *ccm*Fc[a]
	Ubiquinol cytochrome c reductase	*cob*
	Cytochrome c oxidase	*cox1, cox2, cox3*
	Maturases	*mat*R
	Transport membrane protein	*mtt*B
	NADH dehydrogenase	*nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *nad*7, *nad*9
Variable genes	Large subunit of ribosome	*rpl*5, *rpl*10, *rpl*16
	Small subunit of ribosome	*rps*1, *rps*3, *rps*4, *rps*10, *rps*12, rps13, *rps*14, *rps*19
	Succinate dehydrogenase	*sdh*4^2
rRNA genes	Ribosomal RNAs	*rrn*5 (×2), *rrn*8 (×2), *rrn*26 (× 2)^2
tRNA genes	Transfer RNAs	*trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-GCC, *trn*H-GUG, *trn*K-UUU, *trn*M-CAU, *trn*M-CAU, *trn*M-CAU, *trn*N-GUU, *trn*P-UGG, *trn*Q-UUG, *trn*S-GCU , *trn*S-UGA, *trn*W-CCA, *trn*Y-GUA

[a] pseudogene

Table 2 Validation of the homologous recombination in the mitogenome of sweet potato

Repeat name	Repeat 1	Repeat 2	Repeat 3
Identities	94.46	97.43	93.54
Length	253	78	62
Position-1	124,865–125,108	240,244–240,321	117,130–117,191
Position-2	159,977–160,225	245,199–245,275	191,028–191,089
E-value	1.33E-103	1.14E-29	5.39E-18
Type	Direct	Direct	Direct

repeat sequence for PCR amplification experiments (Table S1). Specific primers were designed on each side of the pair of direct repeats present on the main single circular molecule, allowing each of the two PCR products (primers F1 and R1, primers F2 and R2) to span the repeat sequence (Fig. 2A). When the conformation of the recombination was existent, the PCR product (primers F1 and R2, primers F2 and R1) of the exchange of reverse primers was also amplified (Figs. 2B and S2). As all three groups were positive replicates, we used the same strategy to design primers for the experiments, and the final results are displayed in Fig. 2C. All three sets of repeated sequences can produce secondary conformations, which is consistent with the results we obtained based on the long-read analysis.

On the basis of bioinformatics analysis and the results of validation by PCR experiments, we speculated on the potential forms of the mitochondrial genome (Fig. 3). M1 (only one molecule) was the representative structure, and it could form 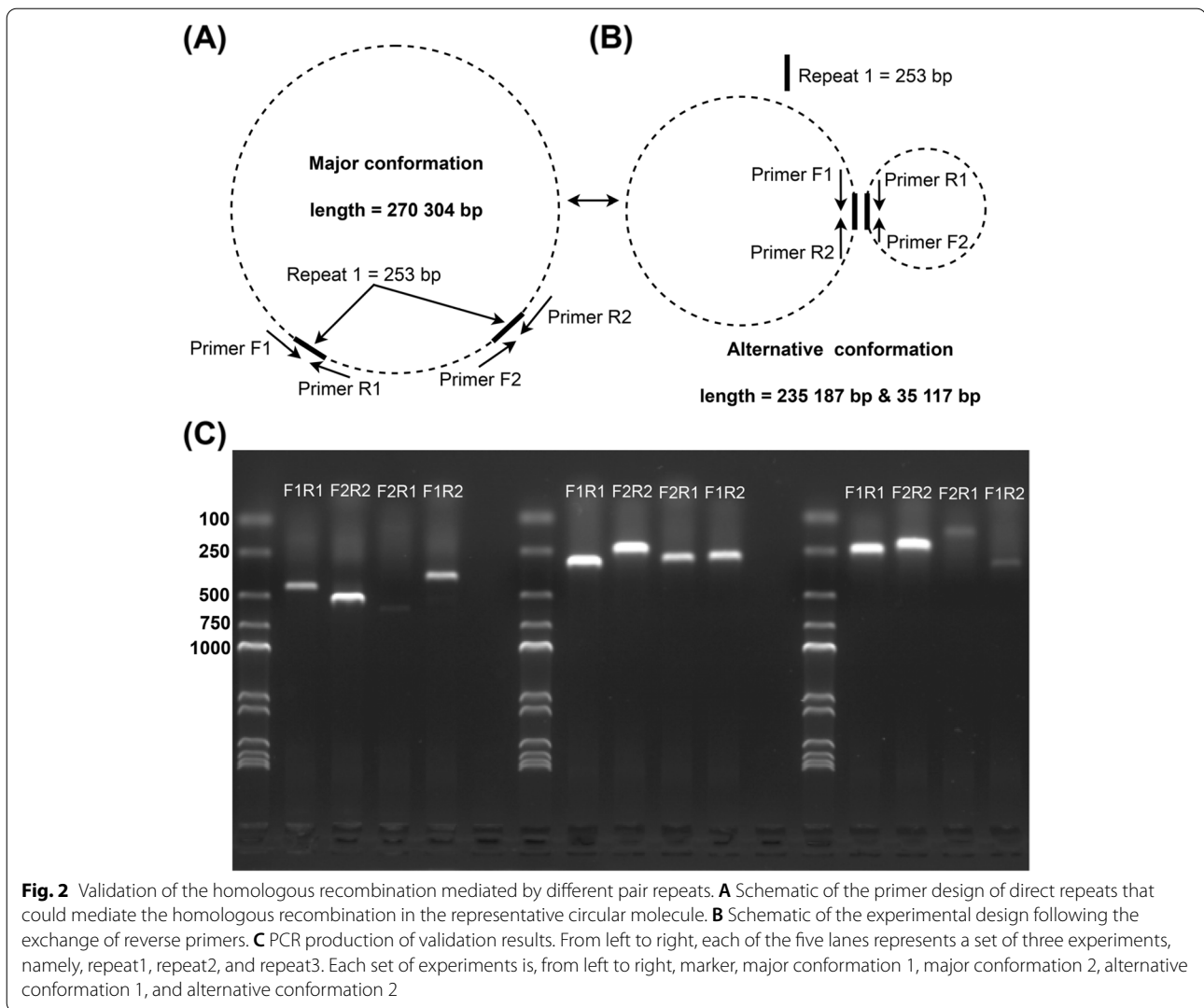three minor conformations M2, M6, and M7 through the three direct repeats R1, R2, and R3, respectively, all of which were bicyclic. In addition, M2 could form three and four cyclic molecular structures through the two repeats of R2 and R3 (Fig. 3).

Repeat elements analysis

Microsatellites (simple sequence repeats [SSRs]) are often used for molecular marker design because of their high polymorphism and codominant inheritance [37]. We used the Misa web server (https://webblast.ipk-gatersleben.de/misa/) to obtain the SSRs in the mitochondrial genome in sweet potato (Table S2, Fig. 4), and we identified 70 SSRs. In the mitochondrial genome, the most frequently occurring SSR was tetranucleotide, accounting for 41.4% SSRs (Table S2, Fig. 4). We also detected dispersed repeats in the mitochondrial genomes of sweet potato (Table S4–5). The dispersed repeats were drawn as colored lines linked at corresponding positions on the genome (Fig. 4). A total of 599 (forward: 303; palindromic: 296) dispersed repeats were found in the mitochondrial genome of sweet potato, and no complement repeats were found. The longest dispersed repeat in mitochondrial genome was palindromic repeat with a length of 915 bp (Table S3, Fig. 4). These repeat sequences have the potential to mediate genomic rearrangements and homologous recombination. In this study, we identified four sets of forward (direct) repeats that could form multiple circular molecules in the mitochondrial genome.

DNA transfer and phylogenetic analysis

The chloroplast genome sequences were integrated into the mitochondrial genome during the evolution. We found 27 homologous DNA fragments in the sweet
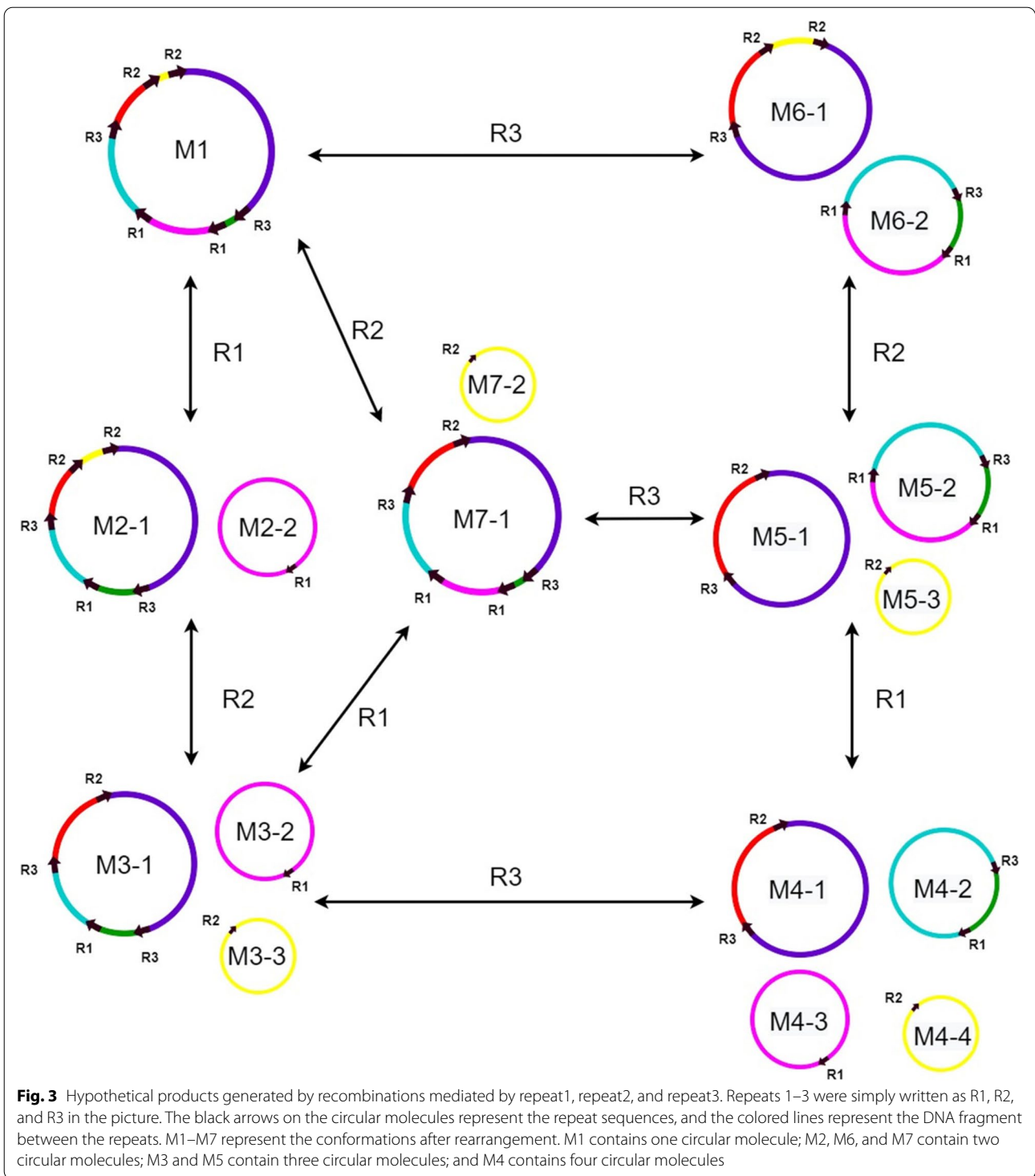
Fig. 2 Validation of the homologous recombination mediated by different pair repeats. **A** Schematic of the primer design of direct repeats that could mediate the homologous recombination in the representative circular molecule. **B** Schematic of the experimental design following the exchange of reverse primers. **C** PCR production of validation results. From left to right, each of the five lanes represents a set of three experiments, namely, repeat1, repeat2, and repeat3. Each set of experiments is, from left to right, marker, major conformation 1, major conformation 2, alternative conformation 1, and alternative conformation 2

potato mitochondrial genome (Table S4, Fig. 5A). The summary insert length is 19,887 bp and accounts for 7.35% of the mitochondrial genome. To further explore the evolutionary relationships of mitochondria in sweet potato, we conducted a phylogenetic analysis of the mitochondria of 16 species of convolvulaceae (*Ipomoea quamoclit* MZ240732, *Ipomoea nil* NC031158, *Ipomoea batatas, Ipomoea aquatica* MZ240730, *Ipomoea biflora* MZ240723, *Argyreia velutina* MZ240724, *Merremia hederacea* MZ240731, *Convolvulus arvensis* BK059236, *Calystegia soldanella* MZ240725, *Evolvulus alsinoides* NC058741, *Dinetus racemosus* MZ240727, *Erycibe obtusifolia* MZ240728, *Cuscuta europaea* BK059238, *Cuscuta epilinum* BK059237, *Cuscuta campestris* BK016277, *Cuscuta japonica* MZ240726). These species have a large structural variation between them. Thus, we used a shared conserved PCG tree building approach. Twenty-four PCG (*atp*1, *atp*4, *atp*8, *ccm*C, *co*b, *cox*1,

*cox*2, *mat*R, *nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *nad*7, *nad*9, *rpl*5, *rpl*16, *rps*3, *rps*4, *rps*12, *rps*13, and *rps*19) were used for phylogenetic analysis (Fig. 5B). All bootstraps exceeded 0.89, indicating the reliability of the inferences from the phylogenetic analysis. According to our analysis, four cuscuta species clustered into one clade and others clustered into one clade. The sweet potato (*I. batatas*) was closely related to the *I. nil* and *I. quamoclit*.

RNA editing events in sweet potato

We predicted RNA editing events for 35 PCGs. A total of 597 RNA editing sites met the requirement of coverage depth > 10 and were distributed among all PCGs (Table S5). *ccm*C, *nad*4, *rpl*16 and *rps*3 genes were edited more than 30 times, and the *rps*3 gene was edited 54 times, which occurred most frequently (Fig. 6A). Fourteen sites had 100% editing efficiency, and 67.8% of the sites had more than 50% editing

Fig. 3 Hypothetical products generated by recombinations mediated by repeat1, repeat2, and repeat3. Repeats 1–3 were simply written as R1, R2, and R3 in the picture. The black arrows on the circular molecules represent the repeat sequences, and the colored lines represent the DNA fragment between the repeats. M1–M7 represent the conformations after rearrangement. M1 contains one circular molecule; M2, M6, and M7 contain two circular molecules; M3 and M5 contain three circular molecules; and M4 contains four circular molecules

efficiency (Fig. 6B). We finally found 11 types of RNA editing in this mitochondrial genome and did not find C to G editing (Fig. 6C). C to U editing was present in all PCGs and has the highest number of occurrences (534 times, 89.4%). G to C and T to A editing were found in the *cox*2 and the *atp*9 gene only once. In

Yang *et al. BMC Plant Biology*   (2022) 22:285

Page 8 of 12
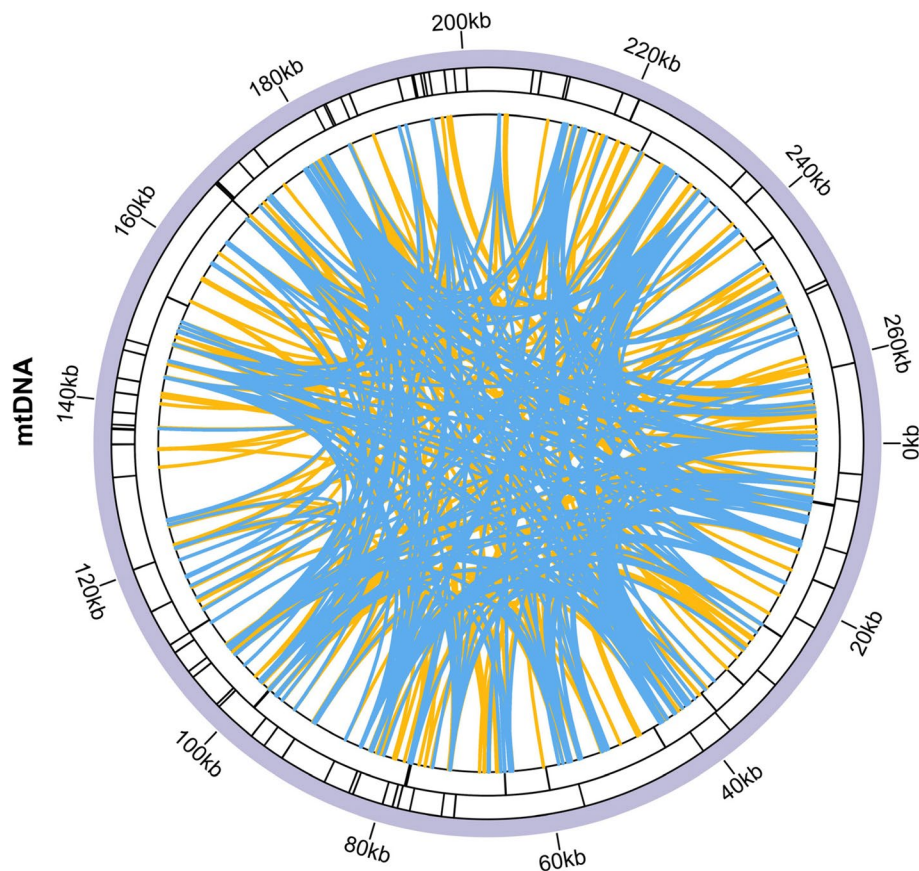
Fig. 4 Repeat analysis of the mitochondrial genome in sweet potato. The inner circle shows the dispersed repeats connected with yellow and blue links. The next circles show the tandem repeats and microsatellite as short bars. The interval scale was 20 kb

summary, RNA editing events occur mainly on C to U and with appreciable efficiency.

Discussion

In this study, the mitochondrial genomes of sweet potato were assembled, and potential substructures were identified. In assembling the mitochondrial genome, we used a graph-based genome that combined the assembly of NGS reads with long reads to assemble the mitochondrial genome of sweet potato (Figure S1). This approach has the advantage of avoiding inconsistencies in the use of NGS data to polish the results of the long-read assembly [38]. Long-read sequencing such as nanopore has a high sequencing error rate and by way of embellishment can lead to assembly results with incorrect bases, especially in protein-coding regions, which can further cause other problems such as stop codons in regions in the middle of PCGs in the mitochondrial genome [39].

Unlike the conserved monocyclic structure of plant chloroplast genomes, seed plant mitochondrial genomes generally have multiple alternative conceptions or minor conformations due to repeat sequences [40, 41]. Jingling Li found a 175 bp direct repeat that can divide the *Scutellaria tsinyunensis* mitochondrial genome into two dependent circular molecules [36]. Shanshan Dong reported that the mitochondrial genome of the *Nymphaea coloratal* may undergo low-frequency recombination under the influence of repeat sequences [28]. Shuaibin Wang discovered the existence of two circular molecules that differed greatly in size in the kiwi fruit mitochondrial genome [42]. The mitochondrial genomes of *Lactuca* were detected with a variety of linear, branched, and circular structures [43]. In summary, plant mitochondria are polymorphic and cannot be represented using only a single circular molecule. Three pairs of repeat sequences found in this study could allow the sweet potato mitochondrial genome to form four separate circular molecules. These phenomena may be due to the specific DNA repair mechanism in the plant mitochondrial genome [44]. We verified only that these cyclic structures existed, but whether those four

Yang *et al. BMC Plant Biology*   (2022) 22:285

Page 9 of 12
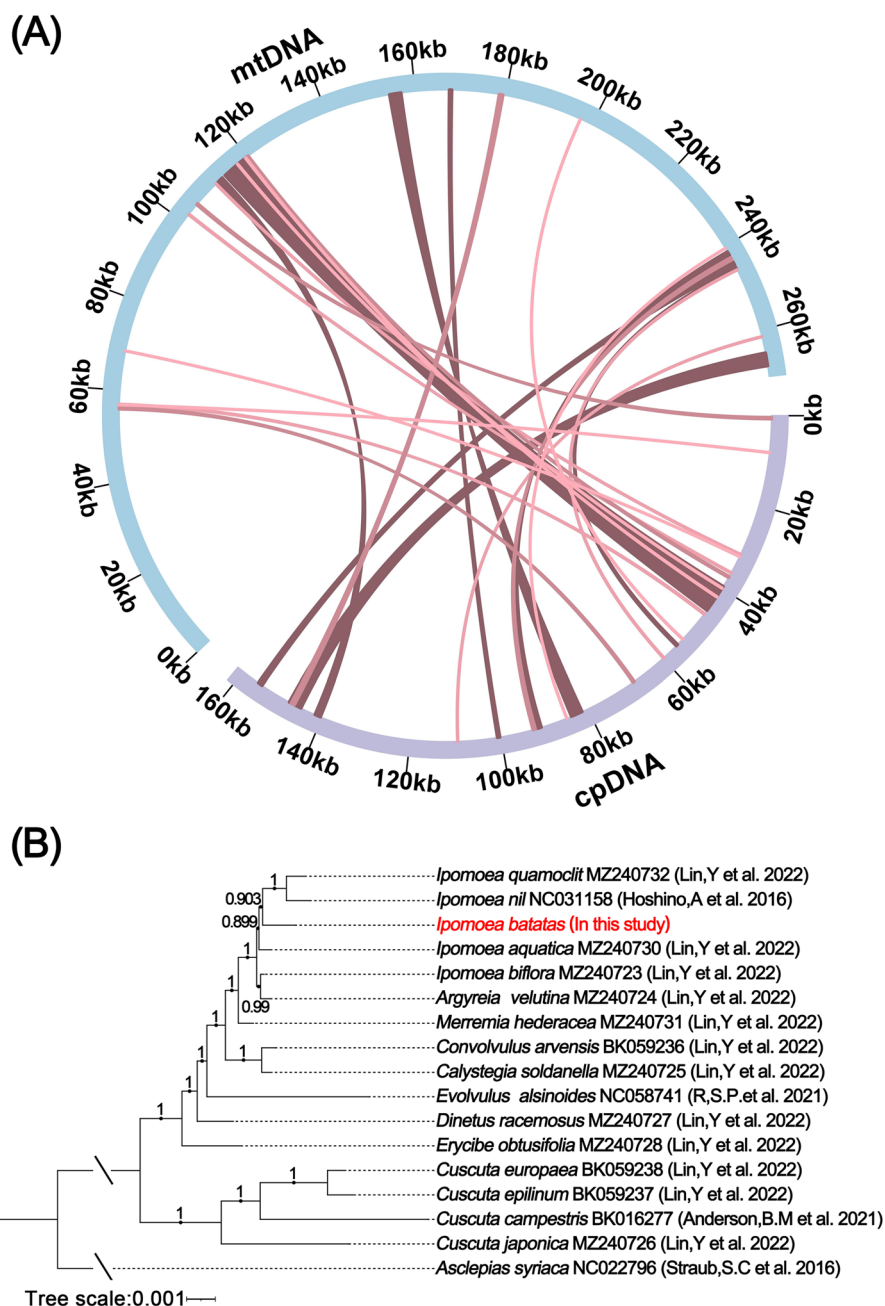
Fig. 5 Homologous fragments and phylogenetic analysis of sweet potato. **A** DNA transfer between the chloroplast and mitochondrial genome. The blue circular segment represents the mitochondrial genome, and the purple circular segment represents the chloroplast genome. **B** Phylogenetic relationships of sweet potato

circular molecules could exist simultaneously requires further investigation.

In addition, the RNA editing events in sweet potato mitochondrial genome were investigated. We found 597 RNA editing sites, which is highly similar to other land plants [45–50]. RNA editing events may affect the start or end points of PCGs. We hypothesized that the *rps*10 gene with start codon ACG appeared because a C to U editing event occurred at the starting position. However, we did not find transcriptome comparison reads to support this phenomenon. According to a previous report, the *cox*1 gene was transcribed through an RNA editing event that allowed ACG to be edited to AUG and was thus used as a starting point in potato
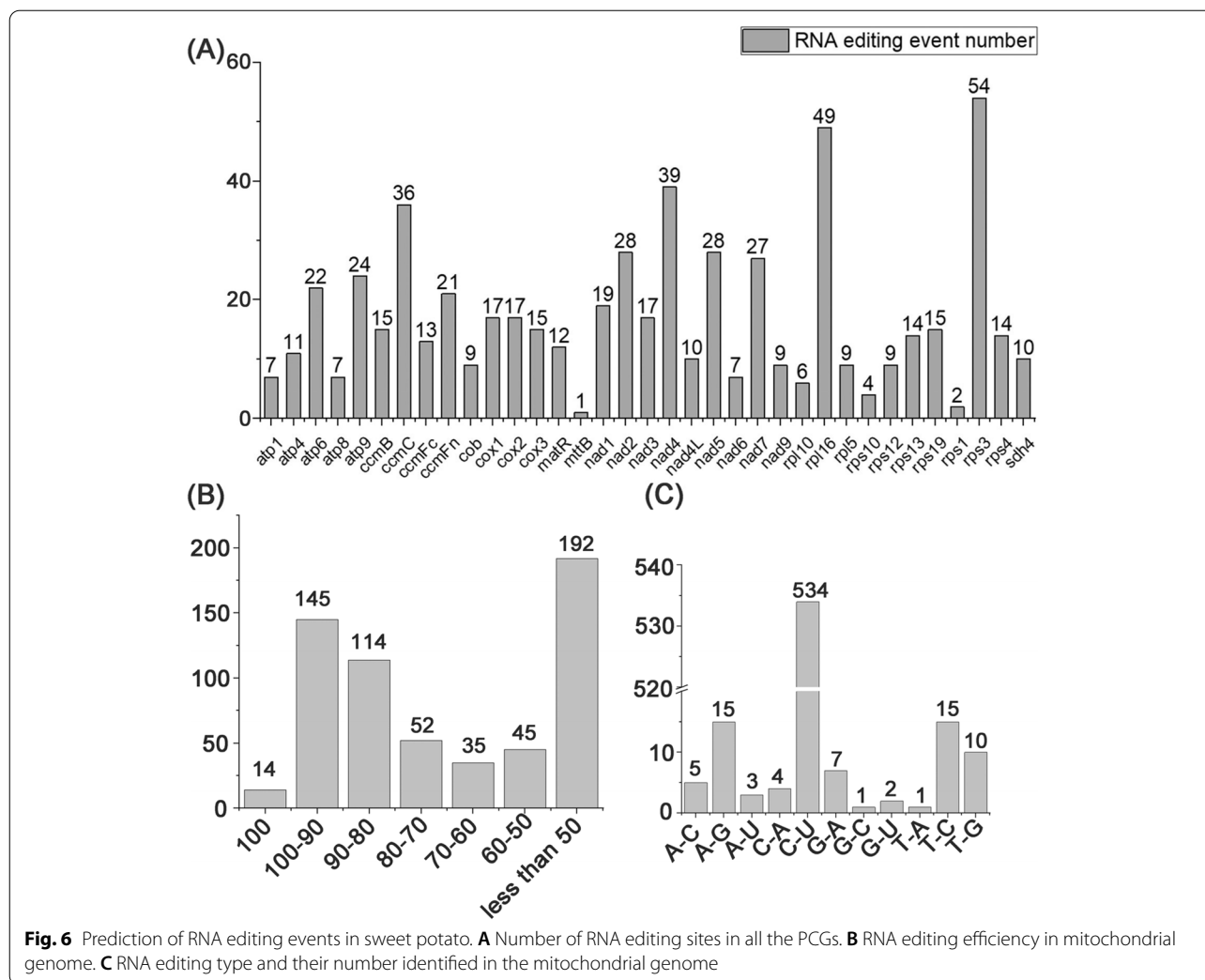
Fig. 6 Prediction of RNA editing events in sweet potato. **A** Number of RNA editing sites in all the PCGs. **B** RNA editing efficiency in mitochondrial genome. **C** RNA editing type and their number identified in the mitochondrial genome

[51]. However, ACG could be used directly as starting points for transcription without editing in organelle genomes [52]. Therefore, whether the starting point of rps10 in sweet potato is edited or not still needs further experimental verification.

Previous studies found that the mitochondrial genomes of the crops were associated with CMS and that the stability of the mitochondrial genome was regulated by nuclear gene expression [53–55]. The sweet potato nuclear genome is homozygous and heterozygous for six ploidy, and varieties can be divided into 15 infertile clusters [13]. Generally, self-crosses of the same variety are incompatible (self-fertility), crosses between varieties within a cluster are incompatible, and crosses between varieties in infertile clusters can bear fruit. The plant trait CMS is determined by the mitochondrial genome and is associated with a pollen sterility phenotype that can be suppressed or counteracted by nuclear genes known as restorer-of-fertility

genes [54]. Dissecting the mitochondrial genome of sweet potato can provide a theoretical basis for CMS breeding in sweet potato.

Conclusion
We successfully assembled the mitochondrial genome of sweet potato by using both nanopore reads and Illumina reads. The mitochondrial genome of sweet potato contains three pairs of direct repeats (253 bp, 78 bp, and 62 bp) which can mediate homologous recombination, and the mitochondrial genome of sweet potato could form a polycyclic structure under the influence of these repeats. Further studies may be conducted in the future to determine whether these circular molecules co-exist and how they relate to CMS breeding.

Yang *et al. BMC Plant Biology*   (2022) 22:285

Page 11 of 12

Supplementary Information

Additional file 1: Table S1. The primers design for repeats mediate the homologous recombination *I. batatas*. **Table S2.** Microsatellite repeats in the mitogenome. **Table S3.** Dispersed repeats in the mitogenome. **Table S4.** The DNA transfer in the mitochondrial genome of *I. batatas*. **Table S5.** The RNA editing events prediction in the *I. batatas*. **Figure S1.** The graph-based mitochondrial genome of *I. batatas*. **Figure S2.**The raw Gel diagram of agarose gel electrophoresis.

Availability of data and materials
The mitogenome sequences supporting the conclusions of this article are available in GenBank (https://www.ncbi.nlm.nih.gov/) with accession numbers: OL699988. The mitochondrial genome also could download from the Figshare platform with the public links (https://figshare.com/s/7c78c9f9ae924d2281ec). The sample has been deposited in the Fujian Agriculture and Forestry University (Fuzhou, China) with accession number IP01. The raw data have been submitted to the SRA database (NGS: SRR17210632; ONT: SRR17210633).

Declarations

Ethics approval and consent to participate
We collected fresh leaf materials of sweet potato for this study. The plant sample was identified by Professor Xuanyang Chen in the Fujian Agriculture and Forestry University. We prepared the voucher specimens and deposited them in the Fujian Agriculture and Forestry University with the accession numbers IP01. The study, including plant samples, complies with relevant institutional, national, and international guidelines and legislation. No specific permits are required for plant collection.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing and conflicting interests.

Author details
[1]Key Laboratory of Crop Biotechnology, Fujian Agriculture and Forestry University, Fujian Province Universities, Fuzhou, China. [2]College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China. [3]Fujian Provincial Key Laboratory of Crop Breeding by Design, Fuzhou, Fujian, China.

References
1. Alam MK. A comprehensive review of sweet potato (Ipomoea batatas [L.] lam): revisiting the associated health benefits. Trends Food Sci Technol. 2021;115:512–29.
2. Motsa NM, Modi AT, Mabhaudhi T. Sweet potato (Ipomoea batatas L.) as a drought tolerant and food security crop. S Afr J Sci. 2015;111(11–12):1–8.
3. Drapal M, Fraser PD. Determination of carotenoids in sweet potato (Ipomoea batatas L., lam) tubers: implications for accurate provitamin a determination in staple sturdy tuber crops. Phytochemistry. 2019;167:112102.
4. Montilla EC, Hillebrand S, Winterhalter P. Anthocyanins in purple sweet potato (Ipomoea batatas L.) varieties. Fruit Vegetable Cereal Sci Biotechnol. 2011;5(2):19–23.
5. Fan G, Han Y, Gu Z, Chen D. Optimizing conditions for anthocyanins extraction from purple sweet potato using response surface methodology (RSM). LWT-Food Sci Technol. 2008;41(1):155–60.
6. Islam S. Sweetpotato (Ipomoea batatas L.) leaf: its potential effect on human health and nutrition. J Food Sci. 2006;71(2):R13–R121.
7. Philpott M, Gould KS, Lim C, Ferguson LR. In situ and in vitro anti-oxidant activity of sweetpotato anthocyanins. J Agric Food Chem. 2004;52(6):1511–3.
8. Li Y, Zhang L, Zhang L, Nawaz G, Zhao C, Zhang J, et al. Exogenous melatonin alleviates browning of fresh-cut sweetpotato by enhancing anti-oxidative process. Sci Hortic. 2022;297:110937.
9. Clark C, Valverde R, Fuentes S, Salazar L, Moyer J: Research for improved management of sweetpotato pests and diseases: cultivar decline. In: I International Conference on Sweetpotato Food and Health for the Future 583: 2001; 2001: 103–112.
10. Hahn S, Isoba JC, Ikotun T. Resistance breeding in root and tuber crops at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. Crop Protection. 1989;8(3):147–68.
11. Sun H, Mei J, Zhao W, Hou W, Zhang Y, Xu T, et al. Phylogenetic analysis of the SQUAMOSA promoter-binding protein-like genes in four Ipomoea species and expression profiling of the IbSPLs during storage root development in sweet potato (Ipomoea batatas). Front Plant Sci. 2021;12:801061.
12. Isobe S, Shirasawa K, Hirakawa H. Challenges to genome sequence dissection in sweetpotato. Breed Sci. 2017:16186.
13. Yang J, Moeinzadeh M, Kuhl H, Helmuth J, Xiao P, Haas S, et al. Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat Plants. 2017;3(9):696–703.
14. Yang Z, Zhu P, Kang H, Liu L, Cao Q, Sun J, et al. High-throughput deep sequencing reveals the important role that microRNAs play in the salt response in sweet potato (Ipomoea batatas L.). BMC Genomics. 2020;21(1):1–16.
15. Xie Z, Wang A, Li H, Yu J, Jiang J, Tang Z, et al. High throughput deep sequencing reveals the important roles of microRNAs during sweetpotato storage at chilling temperature. Sci Rep. 2017;7(1):1–12.
16. Kubo T, Kitazaki K, Matsunaga M, Kagami H, Mikami T. Male sterility-inducing mitochondrial genomes: how do they differ? Crit Rev Plant Sci. 2011;30(4):378–400.
17. Liu H, Cui P, Zhan K, Lin Q, Zhuo G, Guo X, et al. Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. BMC Genomics. 2011;12(1):1–14.
18. Heng S, Wei C, Jing B, Wan Z, Wen J, Yi B, et al. Comparative analysis of mitochondrial genomes between the hau cytoplasmic male sterility (CMS) line and its iso-nuclear maintainer line in Brassica juncea to reveal the origin of the CMS-associated gene orf288. BMC Genomics. 2014;15(1):1–12.
19. 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(1):1–31.
20. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007;23(8):1026–8.
21. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
22. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–W11.

Yang *et al. BMC Plant Biology*   (2022) 22:285

Page 12 of 12

23. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.

24. Stephan G, Pascal L, Ralph B. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nuclc Acids Research. 2019;W1:W59–64.

25. Lewis SE, Searle S, Harris N, Gibson M, Iyer V, Richter J, et al. Apollo: a sequence annotation editor. Genome Biol. 2002;3(12):1–14.

26. Misra S, Harris N. Using Apollo to browse and edit genome annotations. Curr Protoc Bioinformatics. 2005;12(1):9.5. 1–9.5. 28.

27. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.

28. Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, et al. The complete mitochondrial genome of the early flowering plant Nymphaea colorata is highly repetitive with low recombination. BMC Genomics. 2018;19(1):1–12.

29. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.

30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

31. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754–5.

32. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256–9.

33. Sherry S, Xiao C, Durbrow K, Kimelman M, Rodarmer K, Shumway M, et al. Ncbi sra toolkit technology for next generation sequence data. In: Plant and Animal Genome XX Conference (January 14–18, 2012) Plant and Animal Genome: 2012; 2012.

34. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. Bioinformatics. 2010;26(5):589–95.

35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

36. Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, et al. Assembly of the complete mitochondrial genome of an endemic plant, Scutellaria tsinyunensis, revealed the existence of two conformations generated by a repeat-mediated recombination. Planta. 2021;254(2):36.

37. Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet. 2002;30(2):194–200.

38. Wang X, Zhang R, Yun Q, Xu Y, Zhao G, Liu J, et al. Comprehensive analysis of complete mitochondrial genome of Sapindus mukorossi Gaertn.: an important industrial oil tree species in China. Ind Crop Prod. 2021;174:114210.

39. Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. PLoS One. 2021;16(10):e0257521.

40. Ye N, Wang X, Li J, Bi C, Xu Y, Wu D, et al. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant Salix suchowensis. PeerJ. 2017;5:e3148.

41. Cheng Y, He X, Priyadarshani S, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitochondrial genome of Suaeda glauca. BMC Genomics. 2021;22(1):1–15.

42. Wang S, Li D, Yao X, Song Q, Wang Z, Zhang Q, et al. Evolution and diversification of kiwifruit Mitogenomes through extensive whole-genome rearrangement and mosaic loss of intergenic sequences in a highly variable region. Genome Biol Evol. 2019;11(4):1192–206.

43. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15(8):e1008373.

44. Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol. 2013;5(6):1079–86.

45. Wu B, Chen H, Shao J, Zhang H, Wu K, Liu C. Identification of symmetrical RNA editing events in the mitochondria of salvia miltiorrhiza by Strand-specific RNA sequencing. Sci Rep. 2017;7(1):42250.

46. He ZS, Zhu A, Yang JB, Fan W, Li DZ. Organelle Genomes and Transcriptomes of Nymphaea Reveal the Interplay between Intron Splicing and RNA Editing. Int J Mol Sci. 2021;22(18):9842.

47. Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. Proc Natl Acad Sci. 1996;93(16):8766–70.

48. Picardi E, Pesole G. REDItools: high-throughput RNA editing detection made easy. Bioinformatics. 2013;29(14):1813–4.

49. Ichinose M, Sugita M. RNA editing and its molecular mechanism in plant organelles. Genes. 2017;8(1):5.

50. Chen H, Deng L, Jiang Y, Lu P, Yu J. RNA editing sites exist in protein-coding genes in the chloroplast genome of Cycas taitungensis. J Integr Plant Biol. 2011;53(12):961–70.

51. Quiñones V, Zanlungo S, Holuigue L, Litvak S, Jordana X. The cox1 initiation codon is created by RNA editing in potato mitochondria. Plant Physiol. 1995;108(3):1327–8.

52. Zandueta-Criado A, Bock R. Surprising features of plastid ndhD transcripts: addition of non-encoded nucleotides and polysome association of mRNAs with an unedited start codon. Nucleic Acids Res. 2004;32(2):542–50.

53. Sandhu APS, Abdelnoor RV, Mackenzie SA. Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. Proc Natl Acad Sci. 2007;104(6):1766–70.

54. Chase CD. Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. Trends Genet. 2007;23(2):81–90.

55. Horn R, Gupta KJ, Colombo N. Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion. 2014;19:198–205.

Publisher's Note