BMC Genomics

RESEARCH Open Access

Mitogenomes comparison of 3 species of *Asparagus* L shedding light on their functions due to domestication and adaptative evolution

He Wu[1], Wenhua Dongchen[1], Yunbin Li[1], Sylvia E. Brown[1], Shugu Wei[2], Chun Lin[1,3], Zichao Mao[1,3,4*] and Zhengjie Liu[1,3,4*]

Abstract

Background *Asparagus* L., widely distributed in the old world is a genus under Asparagaceae, Asparagales. The species of the genus were mainly used as vegetables, traditional medicines as well as ornamental plants. However, the evolution and functions of mitochondrial (Mt) genomes (mitogenomes) remains largely unknown. In this study, the typical herbal medicine *A. taliensis* and ornamental plant *A. setaceus* were used to assemble and annotate the mitogenomes, and the resulting mitogenomes were further compared with published mitogenome of *A. officinalis* for the analysis of their functions in the context of domestication and adaptative evolution.

Results The mitochondrial genomes of both *A. taliensis* and *A. setaceus* were assembled as complete circular ones. The phylogenetic trees based on conserved protein-coding genes of Mt genomes and whole chloroplast (Cp) genomes showed that, the phylogenetic relationship of the sampled 13 species of *Asparagus* L. were not exactly consistent. The collinear analyses between the nuclear (Nu) and Mt genomes confirmed the existence of mutual horizontal genes transfers (HGTs) between Nu and Mt genomes within these species. Based on RNAseq data, the Mt RNA editing were predicted and *atp1* and *ccmB* RNA editing of *A. taliensis* were further confirmed by DNA sequencing. Simultaneously homologous search found 5 Nu coding gene families including pentatricopeptide-repeats (PPRs) involved in Mt RNA editing. Finally, the Mt genome variations, gene expressions and mutual HGTs between Nu and Mt were detected with correlation to the growth and developmental phenotypes respectively. The results suggest that, both Mt and Nu genomes co-evolved and maintained the Mt organella replication and energy production through TCA and oxidative phosphorylation .

Conclusion The assembled and annotated complete mitogenomes of both *A. taliensis* and *A. setaceus* provide valuable information for their phylogeny and concerted action of Nu and Mt genomes to maintain the energy production system of *Asparagus* L. in the context of domestication and adaptation to environmental niches.

*Correspondence:
Zichao Mao
zmao@ynau.edu.cn
Zhengjie Liu
lzj1022@163.com

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

Highlights

- Mitochondrial genomes of *A. taliensis* and *A. setaceus* were assembled and annotated, and phylogenetic trees of 13 sampled species in the genus Asparagus based on both mitochondrial and chloroplast genomes showed that *A. officinalis*, *A. taliensis* and *A. setaceus* are represented species of *Asparagus* L.
- Independent horizontal gene transfers (HGTs) between the mitochondrial and nuclear genomes were detected with reduction in dioecious species, while Mt RNA editing rate was higher in dioecious species with consistency of their higher copies and higher expression levels of involved nuclear coding RNA editing enzyme genes.
- Detected different gene copies and expression levels of both pathways of tricarboxylic acid cycle(TCA) and oxidative phosphorylation genes indicates the different efficiency of citric acid accumulation and ATP synthesis due to adaptation and/or domestication among species of *A. officinalis*, *A. taliensis* and *A. setaceus*.

Keywords *Asparagus* L., Mitochondrial genome, Phylogenetic analysis, RNA editing, Energy production system

Background

Mitochondria, as energy factories, exists in almost all eukaryotic cells providing energies to power cells by continuously producing adenosine triphosphate (ATP) using both TCA cycle and oxidative phosphorylation pathways to maintain the energy requirements of all biological activities [1]. With the development of DNA sequencing and bioinformatic technologies, both whole genomes of Nu and organelle of many important species were gradually sequenced, assembled and annotated with increasing speed. However, due to lots of repeats accumulation which enlarges the size of Mt (range 0.1 ~ 4 Mb), the number of plant species with assembled Mt genomes are still far less than that of assembled plastid genomes [2]. The Mt genome size of plant species are extremely large compared with that of animals and fungi [3] having a size larger than 100 kb, and in some gymnosperms species even reaching 1–4 Mb [4, 5]. Additionally, most eukaryotic Mt genomes are considered as singular circular genomes, while some higher plants show more complex configurations of Mt (e.g., Y-type, H-type as well as multi-circular structures) [6–8]. With their enlarged size, the non-coding region of plant Mt genomes are much larger than their coding region making the assembly of complete Mt genomes difficult, which normally need the long reads sequence data obtained by ONT or PacBio platforms. Additionally due to adaptive evolution, frequent gene rearrangement and mutual HGTs between Nu and organelle genomes normally occur to make the Mt genomes contain many different originating fragments from Nu and/or plastid genomes [9–11].

Asparagus L. contains more than 200 species distributed across the old world [12, 13] with southern Africa being the center of origin, and southwestern China, regarded as a center of diversity for dioecious species. There are 8 dioecious species including, *A. taliensis* being endemic in Yunnan province of China [14]. Garden asparagus (*A. officinalis*) is cultivated globally as a high value vegetable crop for thousands of years [15]. Other species such as *A. cochinchinensis* (dioecious) [16] and *A. racemosus* (hermaphrodite) [17] are traditional medicinal plants used in China and India respectively. *A. setaceus* and *A. densiflorus* are hermaphrodite species originating from Southern Africa and are used as ornamental plants cultivated and distributed worldwide [18, 19]. *Asparagus* L. species having both hermaphrodite and dioecious species, are traditionally classified into 3 subgenera, in which the hermaphrodite species are grouped into the subgenera of *Protasparagus* and *Myriphyllum*, while the dioecious species are only classed into the subgenus of *Asparagus* [20]. The typical dioecious plant, *A. officinalis* is not only rich in a variety of essential amino acids, vitamins and minerals but also accumulate healthy compounds including steroids and flavonoids having various physiological activities [21]. After continuous cultivation and domestication by humans for thousands of years, compared with other plants in the genus *Asparagus* L., *A. officinalis* grows relatively faster, with higher yields of young stem as harvesting organs cultivated as cash crops. *A. taliensis* is also dioecious and has been used for a long time, but has recently been cultivated as an herbal medicinal plant (by natives of Yunnan) having higher content of steroidal saponins accumulated in its roots system. *A. setaceus* belonging to *Protasparagus* is cultivated as an ornamental plant, mainly growing in gardens or containers as a household plant, with tolerance to shading and relatively weaker resistance to stress, having slow growth and biomass accumulation. The Nu genomes of both *A. officinalis* and *A. setaceus* have been reported [22, 23], and we recently sequenced, assembled and annotated the Nu genome of *A. taliensis* (unpublished data) providing the basis for analysing the evolution of the energy producing system via TCA and oxidative phosphorylation pathways. However, only the Mt genome of *A. officinalis* is currently available [24] limiting the functional analyses of the energy producing system between Nu and Mt genomes of these representative species of *Asparagus* L.

In this study, we assembled the Mt genomes of *A. taliensis* and *A. setaceus*, aiming to (i) comparatively

analyze the assembled Mt genomes of *A. officinalis* and use them to reconstruct the phylogenetic relationships between species of *Asparagus* L; (ii) collinearly analyze the Mt and Nu genomes inferring the mutual HGTs between Nu and Mt genomes within each of the 3 species; (iii) predict and validate possible RNA editing sites of some genes in the Mt genome, and find out the main candidates involved in Nu coding gene families for Mt RNA editing; and (iv) detect the Mt genome variations, differential expression genes (DEGs) and changed HGTs, correlating to metabolism and phenotypes, for analyzing the co-evolution of Nu and Mt genomes.

Materials and methods

Plant material

The green variety "Guelph Millennium" of *A. officinalis* (Aof_G) and the male of *A. taliensis* (Ata_M), were planted in the field of Yunnan Agricultural University. The roots, stems and flowering twigs of Aof_G and Ata_M samples, named Aof_GR, Aof_GS & Aof_GF and Ata_MR, Ata_MS & Ata_MF respectively, were sampled with 3 biological replicates. The Herbarium voucher samples of *A. dauricus*, *A. filicinus*, *A. longifiorus*, *A. neglectus*, *A. oligoclonos*, *A. persicus*, *A. angulofractus*, *A. setaceus*, *A. cochinchinensis*, *A.meioclados* and *A*, *lycopodineus* were obtained from the Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences. The detail informations are listed in sample sheets of Table S1.

DNA sequencing and data processing

The genomic DNA were extracted using the plant genomic DNA kit (TianGen, Beijing, China), while the total RNA which were extracted using the plant RNA-Prep pure plant kit (TianGen, Beijing, China) were reverse transcribed to cDNA using kit of GoScript™ Reverse Transcription Mix, Random Primer (Promega, Beijing China) following the manufacturer's protocols. Genomic DNA or cDNA (~10 μg) was sequenced using the Illumina Genome Analyzer II platform (Illumina, San Diego, CA, USA) with libraries of 150 bp × 2 paired-end (PE) reads. In addition, long reads were generated using the Oxford Nanopore Technologies (ONT) PromethION platform (ONT, Oxford, United Kingdom). All Illumina raw data were evaluated with FastQC [25] then filtered with Fastp [26] to get clean data with default parameters resulted in each clean data of ~6 GB from cDNA for RNAseq analyses, while ~30 G from genomic DNA for organelle genomes assembly respectively. The ~75 GB ONT raw data was checked with longQC (https://github.com/yfukasawa/LongQC) with average Q value of 8.27 which were used for mitogenome assembling. Finally, Illumina and ONT data of *A. setaceus* (with project ID: PRJNA564485) were downloaded from NCBI for mitogenome assembling of *A. setaceus*. All the sequencing data are available on China National Center for Bioinformation (CNCB) with Project IDs of both PRJCA011431 and PRJCA020872. The detail information about these data (including stored website) are listed in the sequencing data sheet of Table S1 and part of "Availability of data and materials" (below).

RNAseq data processing

The transcriptome data (SRR10177391,SRR10186988 and SRR10187001) of *A. setaceus* were downloaded from NCBI. All raw data, both sequenced (*A. officinalis* and *A. taliensis*) or downloaded (*A. setaceus* RNAseq data), were evaluated with FastQC and then filtered with Fastp with default parameters to get clean data. The clean reads were further mapped to the corresponding species reference Nu and Mt genome using hisat2 v2.2.1 [27]. Expression quantification was performed through script of featureCounts [28] to obtain TPM (standardized expression units per million mapped fragments per thousand base exons) matrixes. All gene expression analysis was based on the TPM matrix. The predicted protein sequences of *A. officinalis*, *A. taliensis*, and *A. setaceus* genomes were submitted to eggNOG (http://eggnog-mapper.embl.de/) for functional annotations.

De novo assembly of *A. Setaceus* and *A. taliensis* mt genomes

GSAT [29] were used for Mt genome hybrid assembling with both Illumina and third generation DNA sequencing data. GetOrganelle v.1.7.5.2 [30] was used to assemble Mt draft genomes from the Illumina clean data, when the third generation DNA sequencing data were not available. Racon v1.4.21 [31] was used to self-correct the assembled contig, using Nanopore sequencing data. Tthe entire process was iterated three times, and pilon v1.24 [32] was used to align the Illumina data to the corrected sequence. The entire process was iterated twice. Finally, genome function annotation and mapping were performed using the online source of Geseq (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) with the Mt genome of *A. officinalis* as reference and default parameters.

Phylogenetic analyses

Based on the assembled or downloaded Mt and Cp genomes of the genus *Asparagus* L., 19 Mt complete single copy protein-coding genes sequences *(atp1*, *atp4*, *atp6*, *cox1*, *nad3*, *nad7*, *rps1*, *rps2*, *rps4*, *rps7*, *rps12*, *rps19*, *rpl5*, *rpl16*, *ccmB*, *ccmC*, *ccmFn*, *matR*, *cob)* from 13 species of the genus *Asparagus* L. were obtained by OrthoFinder analyses [33], which were linked together as pseudogene sequence of each species respectively to perform multiple sequence alignments using mafft [34].

After, trimming the gap using trimAI [35], the Mt phylogenetic tree was finally constructed using the maximum likelihood method of RAxML [36] with 1000 bootstrap replicates. Simultaneously based on whole Cp genome, the Cp phylogenetic tree was constructed using RAxML. Both Mt and Cp phylogenetic trees were ultimately beautified with iTOL online website(https://itol.embl.de/).

Collinearity analysis of mt genome among species

Based on the General Feature Format (GFF) annotation file and protein-coding genes sequences, the Mt genome of each species was directly analyzed for collinearity using MCScanX [37] with default parameters, and the results were visualized using the Advance Circos module of TBtools [38].

Analysis of metabolism pathway of mt genome homologous nu genes

Alignment analysis between Mt and Nu genomes were conducted with BLAST v.2.11.0+ [39], the results were then visualized using jcvi (https:/github.com/tanghaibao/jcvi), then their protein sequences of homologous Nu genes and their five upflanking and downflanking genes were extracted. The eggNOG-mapper website (http:/eggnog-mapper.embl.de/) was used for protein functional annotation, the clusterProfiler package v.4.0 [40] was used for kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis and the pheatmap package of R (https://cran.r-project.org/web/packages/pheatmap) was used to visualize the genes expression patterns.

Prediction and verification of RNA editing sites

The coding gene sequence of each Mt genome was submit to PREP-mt online prediction software for RNA editing prediction [41]. To eliminate the impact of mutation (as much as possible), the DNA sequence and transcriptome data of the species were aligned with the predicted results of PREP-mt using BWA v0.7.17 [42] and hisat2 v2.2.1 respectively. The results were then viewed using Integrative Genomics Viewer (https://igv.org/). Finally, *atp1* and *ccmB* genes of *A. taliensis* were selected for amplification using templates of both DNA and cDNA, respectively, to verify the prediction results of RNA editing with PCR primer listed in Table S7.

Nuclear RNA editing enzymes gene family analyses

The hmm model corresponding to PentatricoPeptide-Repeat (PPR), Multiple Organellar RNA Editing Factors (MORF), Organelle RNA Recognition Motif (ORRM), Organelle Zinc-finger (OZ) and ProtoporPhyrinogen IX Oxidase 1 (PPO1) gene family were obtained from Pfam (https://pfam.xfam.org/) and screened with hmmsearch v.3.3.1 (filtered by e-value < 1e-3, http://hmmer.org/), while the protein sequences of RNA editing gene families with known functions in *Arabidopsis* and other plants were downloaded from NCBI to screen the candidates with BLAST v.2.11.0+. MUSCLE v3.8.1551 was used to perform multiple sequence alignments between the protein sequences and the functional known proteins in *Arabidopsis* [43]. RAxML was used to construct a phylogenetic tree with 1000 bootstrap replicates. The protein sequences under the same branch were obtained to conduct motif analysis using MEME Suite v5.5.1 (https://meme-suite.org/meme), while the localization of signal peptide prediction was performed using Predotar v.1.0.4 [44] (https://urgi.versailles.inra.fr/predotar/) on all nuclear proteins. The eggNOG was used to annotate the gene function, and the expression was visualized using pheatmap package of R.

Genes related to mt function and their expression analyses

All genes related to Mt reproduction, oxidative phosphorylation and citrate cycle were selected from the KEGG pathway of *Arabidopis thalianta* (https://www.kegg.jp/kegg-bin/show_organism?menu_type=gene_catalogs&org=ath), *Oryza sativa* ssp. *Japonica* (https://www.kegg.jp/kegg-bin/show_organism?org=osa) and *A. officinalis* (https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=aof) to obtain their respective proteins ID. Then the protein sequences of both Mt and Nu genomes of *A.thalianta*, *O. sativa* and *A. officinalis* were obtained with seqkit (https://github.com/shenwei356/seqkit) and were used as a database for BLASTP analyses. The Nu and Mt genomes coding proteins of *A. officinalis*, *A. taliensis* and *A. setaceus* were used as queries to screen candidate homologous genes respectively with critical selection standards of identity rate > 0.4, coverage rate > 0.8 and evalue < 1e-6. The gene list related to Mt reproduction, oxidative phosphorylation and citrate cycle of *A. officinalis*, *A. taliensis* and *A. setaceus* were obtained for comparison and gene expression analyses among the three species.

Result and analysis

Mitogenomes structure and phylogenetic analyses

The complete circular Mt genomes of *A. taliensis* (with a total 512,823 bp) and *A. setaceus* (521,341 bp), which were assembled using both Illumina and ONT data (Fig. 1A and C, Table S2), had a size 3 ~ 4-time that of their assembled complete circular Cp genomes (Fig. 1B and D). The assembled Mt genomes of *A. taliensis* and *A. setaceus* were annotated with 53 and 52 genes respectively, which is similar to the 53 genes reported in the Mt genome of *A. officinalis* (Table S3, Fig. 1A and C). The additional 10 sampled species assembled with partial Mt genomes contained 39 annotated genes excluding tRNA genes, but had all complete circular Cp genomes

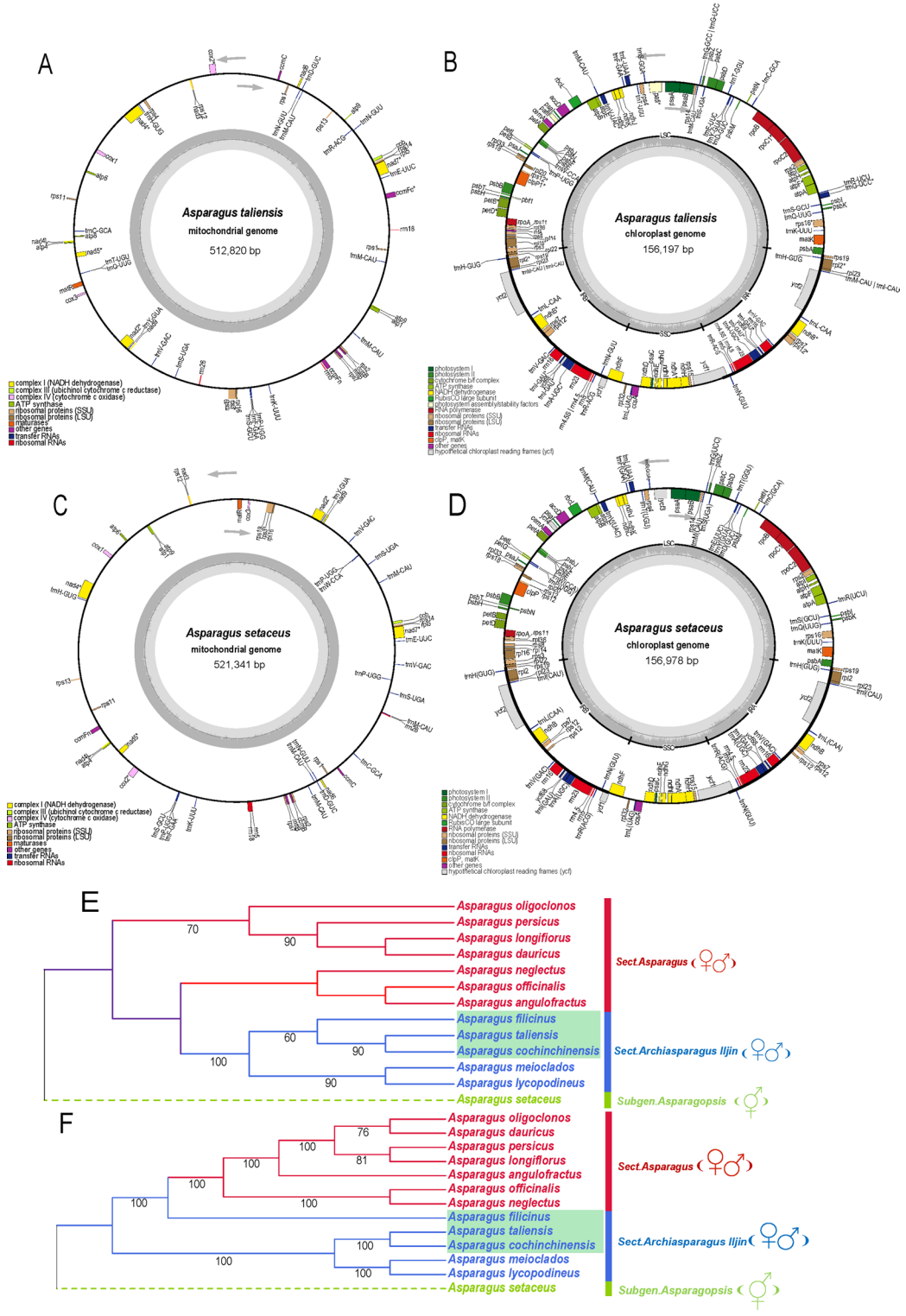

Fig. 1 Mitogenome structure and phylogenetic analyses of *Asparagus L*; **A**, **C**: Mt genomes of *A. taliensis* and *A. setaceus* respectively; **B**, **D**: Cp genomes of *A .taliensis* and *A. setaceus respectively*; **E**, **F**:phylogenetic trees of the Mt and Cp genomes, bootstrap values < 50% not display in branches, classification differences between Mt and Cp in the phylogenetic trees are highlighted with light green while the numbers on the phylogenetic trees represents the bootstrap value of that branch

assembled from Illumina sequencing data using GetOrganelle (Fig. 1E, Table S2). The GC contents of the assembled Mt genomes of *Asparagus* L. species ranged between 46.58%~46.62%. Choosing the hermaphrodite species *A. setaceus* as the outgroup, the phylogenetic trees were constructed using the available complete conserved 19 protein-coding genes of Mt genomes (*i.e. atp1*, *atp4*, *atp6*, *cox1*, *nad3*, *nad7*, *rps1*, *rps2*, *rps4*, *rps7*, *rps12*, *rps19*, *rpl5*, *rpl16*, *ccmB*, *ccmC*, *ccmFn*, *matR*, *cob*) and their whole Cp genomes. The Mt phylogenetic tree showed that, the sampled dioecious species *of Asparagus* L. were grouped into 2 major clades, one consisting of 4 species (i.e. *A. dauricus*, *A. longifiorus*, *A. persicus*, *A. oligoclonos*), and the other consisting of all other dioecious species, in which, *A. officinalis* and *A. taliensis* were grouped into two independent smaller clades (Fig. 1E). The Cp phylogenetic trees also showed that all dioecious species were grouped into 2 major clades, one including *A. lycopodineus*, *A. meioclados*, *A. taliensis* and *A. cochinchinensis*, while the other contained the left dioecious species. In this second clade *A. filicinus* was an independent group which was sister to that consisting of the other dioecious species including *A. officinalis* (Fig. 1F). According to the traditional taxonomy, *A. setaceus* belonged to the subgenus *Asparagopsis* of *Asparagus* (labeled as green) is a hermaphrodite species, while all dioecious species are classified into the subgenus *Asparagus* with 2 branches of sect. *Archiasparagus* lljin (labeled as light blue) and sect. *Asparagus* (labeled as red). Comparison of the Mt and Cp genomes phylogenetic trees showed that the limited sampled species of sect. *Archiasparagus* lljin were grouped into a single clade, which was sister to the clade belonging to sect. *Asparagus* (*A. neglectus*, *A. officinalis* and *A. angulofractus*). However, the phylogenetic tree based on Cp genomes showed that sect. *Archiasparagus* lljin was paraphyletic, while sect. *Asparagus* was monophyletic. It is interesting to note that species of *A. filicinus* which belongs to sect. *Archiasparagus* lljin was sister with all species of sect. *Asparagus* as an independent smaller clade in Cp tree. The Mt phylogenetic trees also showed that *A. taliensis* and *A. filicinus* within sect. *Archiasparagus* lljin grouped into a group which was sister the clade consisting of *A. meioclados* and *A. lycopodineus*, while in the phylogenetic tree of Cp genomes, *A. taliensis* and *A. filicinus* were not merged into one clade. Based on the Mt tree, *A. officinalis*, *A. angulofractus* and *A. neglectus* were grouped in a sister clade to that of sect. *Archiasparagus* lljin. However based on Cp tree, only *A. officinalis* and *A. neglectus* were grouped into a smaller clade besides all other species of the sect. *Asparagus.* Both Cp and Mt trees showed that *A. setaceus*, *A. taliensis* and *A. officinalis* were representing the hermaphrodite and dioecious species of the sect. *Archiasparagus* and sect. *Asparagus*. Simultaneously these 3 species were representative species of ornamental, herbal medicine and vegetable of *Asparagus* L., respectively. Analyzing the co-evolution and adaptation of genes in both Nu and Mt genomes related to the energy producing systems via TCA and oxidative phosphorylation pathway is important for both basic and applied researches of *Asparagus* L.

Collinear analyses of mt genomes among the 3 species

The MCScanX was used for collinear analyses, and the results showed that, the three Mt genomes have similar structures and collinear regions (blocks) (Figure. S1) inferring the similar functions of Mts among the 3 species. In detail, the Mt genome of *A. setaceus* (ASMT) have 14 collinear blocks with *A. taliensis* Mt genome (ATMT) and *A. officinalis* Mt genome (AOMT) respectively, while ATMT and AOMT had 15 collinear blocks. Further analyses of these block containing genes related to energy production by TCA and oxidative phosphorylation in AOMT showed that *Aof-cox2*, *Aof-ccmFc*, *Aof-atp8*, *Aof-cox3 and Aof-nad1* were specific to AOMT and ATMT, while genes: *Aof-cox1*, *Aof-atp6*, *Aof-nad9*, *Aof-ccmC* were specific to AOMT and ASMT. Comparing ATMT with AOMT and ASMT respectively, the genes: *Ata-atp1*, *Ata-ccmFn*, *Ata-ccmB* were specific to ATMT and AOMT, while genes: *Ata-nad4*, *Ata-nad6*, *Ata-cox1*, *Ata-cox3* and *Ata-atp4* were specific to ATMT and ASMT. Similarly, the genes for direct energy production: *Ase-atp1*, *Ase-atp9*, *Ase-atp8*, *Ase-ccmFn*, *Ase-ccmB* were specific to ASMT and AOMT, while *Ase-cox2*, *Ase-cox3*, *Ase-nad4* were specific to ASMT and AOMT. It is interesting to note that *ccmFc* (which are encoding proteins for assembly of heme with c-type apocytochromes [45]) only existed in dioecious species of both *A. officinalis* and *A.taliensis* but was not detected in the Mt of hermaphodite species *A. setaceus.* However, 3 *ccmFc* homologous genes (i.e., *Ase-ccmC, Ase-ccmFn* and *Ase-ccmB*), and 3 additional homologous genes in both ATMT (i.e., *Ata-ccmC*, *Ata-ccmB* and *Ata-ccmFn*) and AOMT (i.e., *Aof-ccmC*, *Aof-ccmFn* and *Aof-ccmB*) were detected in the three species (Table S3). The results also showed inversions or rearragements of gene clusters between Mt genomes among species, for example, ATMT and ASMT had a similar four genes cluster of *cob-nad7*, while they differ from that of the AOMT block (Figure S1). These different blocks and rearrangments among the mitogenomes of the 3 species may be related to the different energy production requirements for survival and adaptation to their environmental niches.

Collinear analyses between mt and nu genomes among the 3 species

For detecting mutual HGTs between Mt and Nu genomes among the 3 species of *Asparagus* L., the collinear region

analyses between the Nu and Mt were conducted. It was found that ASMT had 6 collinear blocks with its Nu genome, in which 4 blocks were found in chromosome 01 (AseChr01, same as below) and 1 collinear block in AseChr03 of *A. setaceus.* AOMT had 1 block in AofChr01, 2 blocks in AofChr08 and 1 block in AofChr09 of *A. officinalis* Nu genome, while ATMT had 2 blocks in AtaChr04 and 2 blocks in AtaChr05 of *A. taliensis* (Figure S2 A, B and C). The Nu genomes of the 3 species also had collinear regions of chromosomes in which Chr01 ~ Chr10 *of A. officinalis* were more homologous with Chr01 ~ Chr10 of *A. taliensis*, while the fragments of Mt genome integrated chromosomes (MGICs) (i.e. AseChr01 and AseChr03) were more homologous with Chr04 and Chr03 of both *A. officinalis* and *A. taliensis* respectively (Figure S2 D). It is interesting to note that, only the MGICs in AtaChr04 (*A. taliensis*) had the integrated gene *rps7* identical to the *rps7* in AseChr01(*A. setaceus*), while all HGTs between Nu and Mt genomes among the 3 species were found to be independent. The results also showed that the hermaphrodite species *A. setaceus*, which is phylogenetically closer to the dioecious common ancestor of both *A. taliensis* and *A. officinalis* (Fig. 1E), had more HGTs between Mt and Nu genomes, while the dioecious species *A. taliensis* and *A. officinalis* were found to have independent but reduced HGTs which may be due the evolution and/or domestication under different contexts of energy requirement depending on environmental niches. These collinear blocks containing genes in Nu genomes of each species, including 5 genes each flanked to its 5' (up) and 3'(down) blocks, were combined for further KEGG enrichment analyses. The enrichment analyses showed that these genes were mainly classified into pathways related to Mt functions (e.g., redox metabolism, oxidative phosphorylation, ribosome biosynthesis, assemble and replication of Mt as well as thermogenesis), the other genes corresponding to multiple specific metabolic or signaling pathways. It is to note that 4 enriched genes including 2 *CYP450s* (*Ata04G040340* and *Ata05G023230*) and 2 transcriptional factors (*Ata05G011840* and *Ata04G040190*) of *A. taliensis* were involved in biosynthesis of secondary metabolites (e.g., isoflavonoids and steroids). These enriched genes were found in *A. taliensis* which was recently cultivated as a medicinal plant that has accumulated more secondary metabolites such as steroidal saponins and isoflavonoids (Fig. 2, Figure S2 and Table S4). It is also interesting to note that, 1/3 of the 24 genes in *A. officinalis* (i.e., *AofChr08.983*, *AofChr08.970*, *AofChr08.981*, *AofChr03.1816*, *AofChr03.1814*, *AofChr03.1819*, *AofChr09.700* and *AofChr09.702*) were annotated to relate to the metabolic and signal transduction pathway, and included 6 genes involved in plant growth and development of plant organs (e.g. Mitogen-activated protein kinase (MAPK) as well as signaling pathway). As for *A. setaceus*, excluding the Mt function related genes, the richen pathway genes were found to mainly relate to purine metabolism and phytohormone biosynthesis and signaling. Interestingly 1 gene of this species (*Ase03G2815*) was annotated for Ultra Violet (UV) damage repair like gene involved in DNA repair which may contribute to *A.setaceus* occasionally grows under high light intensity or direct sunlight conditions repairing DNA damage caused by UV radiation.

After comparing the expression levels of both Mt genes and their corresponding homologous Nu genes within species (Figure S2), it was found that some Nu genome genes for Mt function had relatively low expression levels, while the corresponding homologous genes in Mts exhibited relatively high expressions, so that the average overall expression level tends to stabilize among species.

RNA editing of mt genes and their involved enzymes coding gene families of nu genome

The RNA editing sites of Mt RNA were predicted and the results showed that 36 out of 39 coding Mt genes in AOMT and ATMT had RNA editing with different average sites per gene (SPG) of 15.7 and 14.7 respectively, while 35 out of 38 coding Mt genes in ASMT had RNA editing with average SPG of 15.1 (Table S5). The results also showed 11 genes (i.e., *atp1*, *atp9*, *ccmFc*, *cox1*, *cox2*, *nad1*, *nad2*, *nad5*, *nad6*, *rps4* and *rps7*) with different RNA editing sites among species (Table S5). The RNA editing was mainly classed from C to U editing type, in which the editing sites mainly occurred in the second nucleotide (Nt) of the codon, followed by the first Nt of codon and without detecting the third Nt of the codon (Table S5). Two mitochondrial genes of both *atp1* and *ccmB* were detected to have 11 and 35 editing sites in both *A. taliensis* and *A. officinalis*, while *A. setaceus* had 2 and 35 editing sites (Table S5 & Table S6). To check the results of RNA editing sites, the *Ata_atp1* and *Ata_ccmB* of *A. taliensis* were sequenced for both Mt genomic DNA and RNA derived cDNA. CDS of *Ata_atp1* in 1168, 1178 &1262 sites and CDS of *Ata_ccmB* in 313, 338, 392, 406, 424 & 428 sites were checked to conduct the C to U editing (Fig. 3). For predicting the possible gene families of Nu genome involved in Mt RNA editing enzymes coding, the genes of 5 families which were reported to be involved in RNA editing were used as queries to search homologous candidates of RNA editing gene families of Nu genomes among the species. The following critical standards were used: (i) coding proteins having more than 35% homology and the length of candidate coding proteins have at least more than the average length of the functionally confirmed Mt RNA editing proteins of pentatricopeptide-repeats (PPRs) [46], multiple organellar RNA editing factors (MORFs) [47], organella

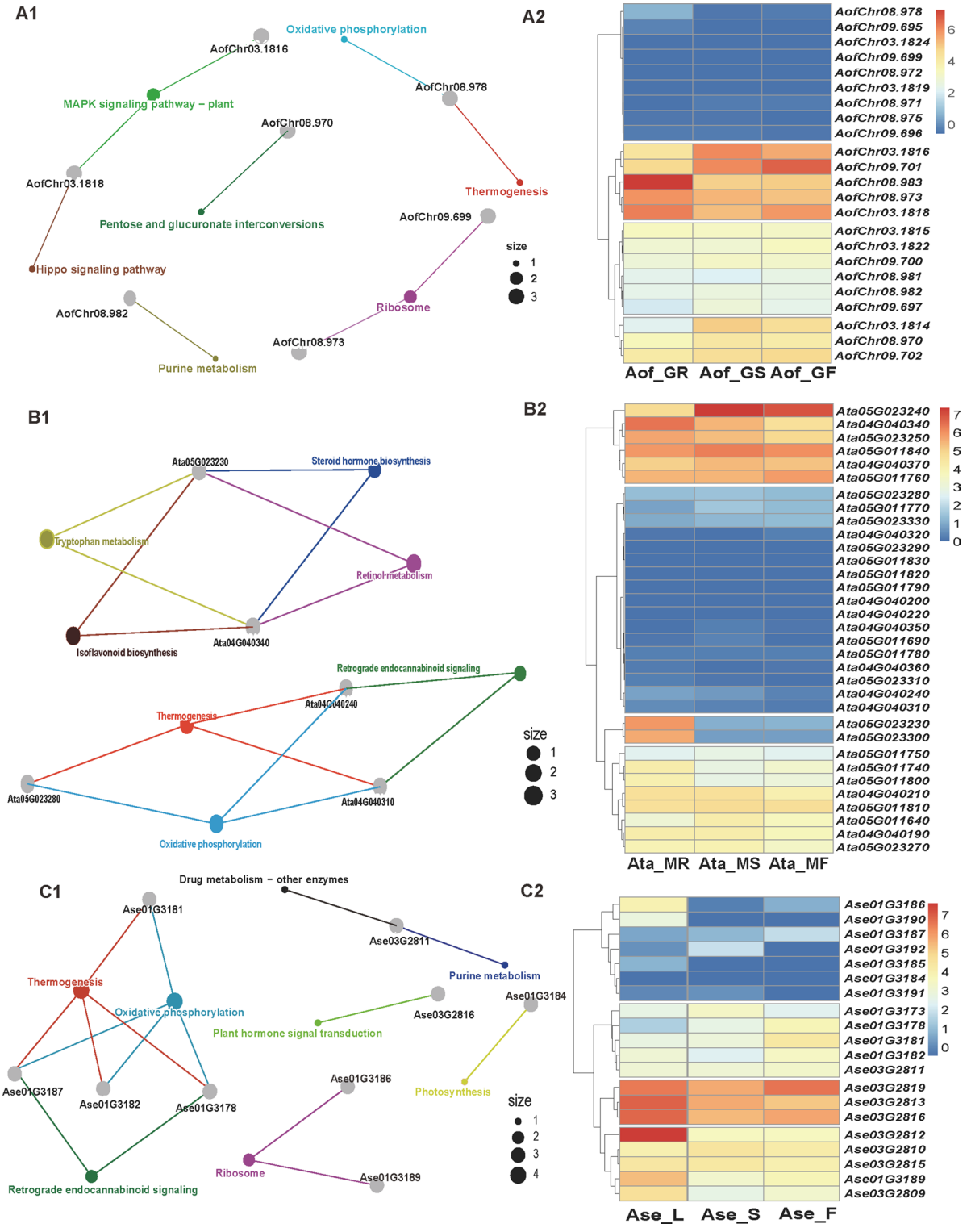

Fig. 2 Metabolic pathways and expressions of partial Nu genes in *Asparagus L*;**A1-C1**:partial KEGG pathway map of *A. officinalis*, *A. taliensis and A. setaceus* collinear Nu genes with their five up and downflanking genes respectively;**A2-C2**: collinear Nu genes heatmap expressions of *A. officinalis*, *A. taliensis and A. setaceus* with their five up and downflanking genes respectively. Where Aof_GR represent green root of *A. officinalis*, Aof_GS represent green stem of *A. officinalis*, Aof_GF represent green flowers of *A. officinalis*, Ata_MR represent wildtype male roots of *A. taliensis*, Ata_MS represent wildtype male stems of *A.taliensis*, Ata_MF represent wildtype male flowers of *A. taliensis*,Ase_L represents leaves of *A. setaceus*, Ase_S represents stems of *A. setaceus* and Ase_F represents flowers of *A. setaceus*

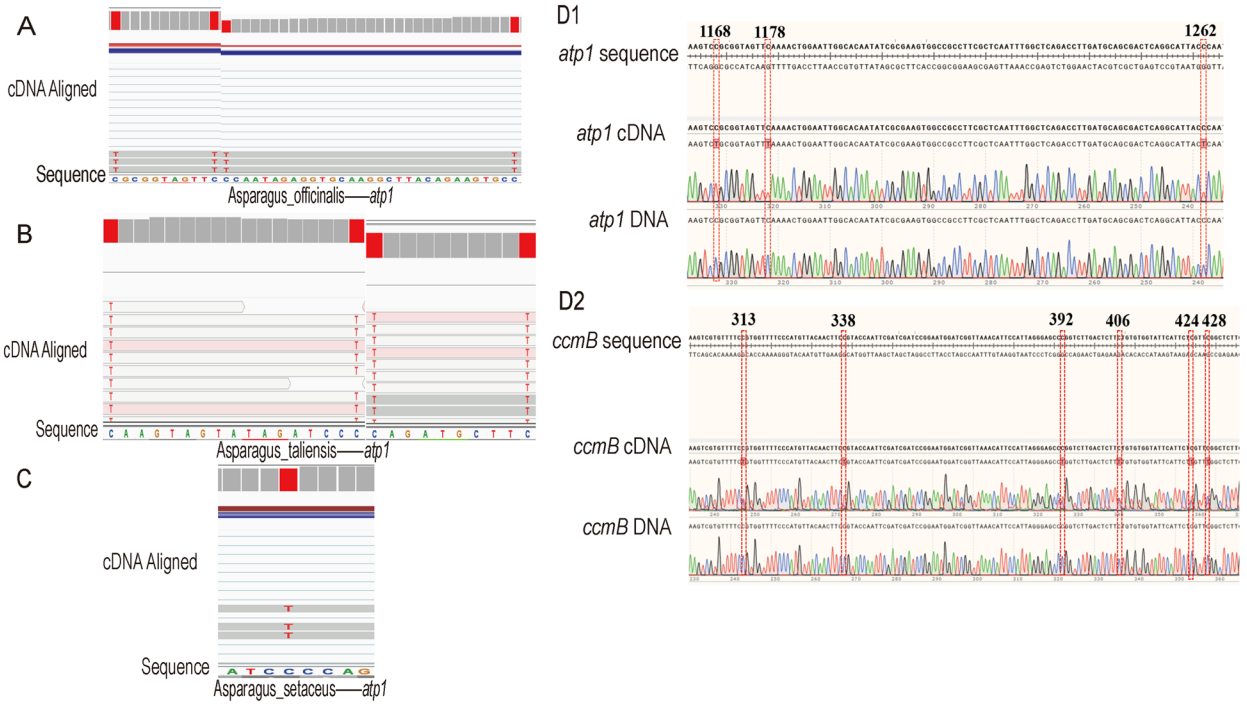

Fig. 3 RNA editing prediction and validation;**A**-**C**: RNA editing prediction of *A. officinalis*, *A. taliensis* and *A. setaceus atp1* gene, the black border indicates that there may be RNA editing at the site of cytosine (C), which transforms into uracil (U, T represent represent their respective positions) during transcription into mRNA;**D1**-**D2**: PCR product sequencing and comparison of *A.taliensis atp1* and *ccmB*, where the red dashed boxes represent RNA editing site confirmed by PCR amplification while the numbers represent the RNA editing site position in the whole sequence

RNA recognition motifs (ORRMs) [48], organella zinc-fingers(OZs) [49] and protoporphyrinogen IX oxidase 1(PPO1) [50], (ii) the candidate genes have expressions (at least 1 TPM) of the sampled organs among species, and (iii) the candidate proteins contains Mt location signals peptide. The results showed that 56, 67 & 60 *PPRs*; 2, 7 & 4 *MORFs* and 3, 1 & 2 *ORRMs* were found in Nu genomes of *A. officinalis*, *A. taliensis* and *A. setaceus* respectively. The results also showed that 2 *OZs* were found in both *A. officinalis* and *A. taliensis* but not in *A. setaceus*, and only 1 candidate *PPO1* was detected in *A. taliensis* (Fig. 4, Table S8). The obtained RNA editing enzymes were used for further conserved motifs analyses and the results showed that all PPRs, MORFs, OZs and PPO1s had 3 consistent conserved motifs in all three species (Fig. 4A, B, D & E); MORFs in both *A. officinalis* and *A. setaceus* had 2 conserved motifs while only one copy of MORFs (Ata04G034700) with one conserved motif in *A. taliensis* was detected (Fig. 4C). These results infer that MORF (Ata04G034700) in *A. taliensis* may be pseudogenes which encode a partial protein sequence with lost or neo-functionalization compared to *A. officinalis* and *A. setaceus*. Further gene expressions of *PPRs* showed that, the *PPRs* of *A. officinalis* (*AofChr05.1937*, *AofChr07.869*, *AofChr04.9*, *AofChr01.3520* and *AofChr01.2935*), *A. taliensis* (*Ata02G024190*, *Ata03G001970*, *Ata08G001000*, and *Ata10G006670*), and *A. setaceus (e.g., Ase02G3097*, *Ase03G1206*, *Ase03G1835*, *Ase05G2973 and Ase08G0092*) were detected with relatively higher expression levels in all sampled organs. Thus, these genes may be key participants in the editing of Mt RNAs including the RNAs of *atp1* and *ccmB*, even though their molecular mechanism still remains unknown.

The functions of Nu and Mt genomes for replication and maintainability of mt organelles, TCA and oxidative phosphorylation

All genes of KEGG pathways involved in replication and maintainability of Mt organelle, TCA and oxidative phosphorylation in *Arabidopsis thaliana*, *Oryza sativa* ssp. *Japonica* and *A. officinalis*, were used as database for the homologous search of genes in the 3 species of *Asparagus* L. using BLAST. The results indicate that the gene numbers related to Mt biosynthesis and assembly in *A. setaceus* (with total 196) were less than both counterparts of *A. officinalis* (207) and *A. taliensis* (258) (Table S9). ASMT, AOMT and ATMT encoded 35, 36 and 36 proteins, respectively, which had expressions at least in 1 sampled organ (Table S9). The major expansion or contraction of gene or gene families were detected with expressions in Nu genome encoded with 80 mtDNA quality control factors (MTQFs) and 91 mtDNA replication factors (MTRFs) in *A. officinalis*, 93 MTQFs and 129 MTRFs in *A. taliensis*, and 84 MTQFs and 77 MTRFs in

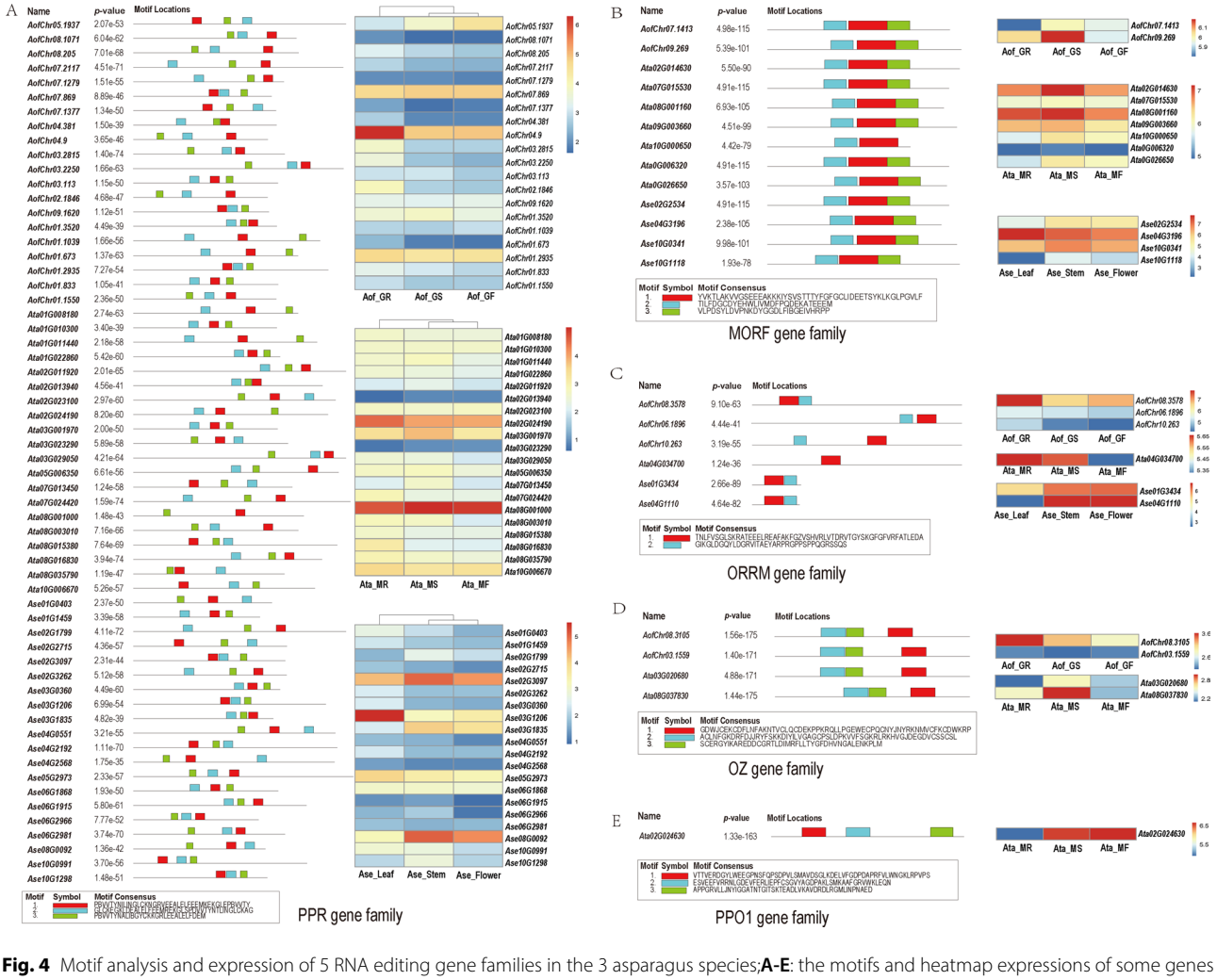

Fig. 4 Motif analysis and expression of 5 RNA editing gene families in the 3 asparagus species;**A-E**: the motifs and heatmap expressions of some genes in the PPR, MORF, ORRM, OZ and PPO1 families of *A. officinalis* (Aof), *A.taliensis*(Ata)and *A.setaceus*(Ase)

A. setaceus. There were more expansions of MTRFs genes in dioecious species (*A. officinalis* and *A. taliensis*) than the counterparts of hermaphrodite species, suggesting higher copy numbers of Mt organelles in cells of dioecious species than hermaphrodite species. It is interesting to note that the *atp1* and *atp9* for coding subunit of ATP synthase complex in Mt of *A. setaceus* had duplicated copies in its Nu and Mt genome respectively, while, *atp1* and *atp4* (subunit of ATP synthase complex) had duplication partial gene fragment (*atp1*) and complete copies (*atp4*) in Nu and Mt genome of *A. taliensis* (Figure S2 and Table S9). However, no ATP synthase subunit coding genes was detected in Nu genome of *A. officinalis.* Additionally, as described, *atp4* of *A. taliensis* were detected with no expression in both Mt and Nu genomes, indicating that, the evolutionary context allowed the progressive removal of copies of ATP synthase complex enzyme genes either in Nu or Mt genome of *Asparagus* L. Further analyses of gene expressions of MTQFs and MTRFs based on RNAseq data showed that a total of 180 genes had differential expressions between the roots (Rs), stems (Ss) and flowering twigs (Fs) of *A. officinalis* and *A. taliensis*, in which 44 differential expression genes (DEGs) were common in all organs, while, 13, 9 & 20 DEGs were specific to Rs, Ss and Fs respectively (Figure S3). Among them, 21 differentially expressed genes were upregulated in *A. officinalis* (Figure S4). From the expression level it can be seen that the overall expression level of AOMT replication factor genes was higher than that of *A. taliensis* and *A. setaceus*, indicating that the copy number of mitochondria in a single cell of *A. officinalis* may be higher. Therefore, increasing the number of mitochondria in a cell can improve energy metabolic efficiency to meet *A. officinalis* growth and developmental needs. The DEGs patterns of these organs between *A. officinalis* and *A. taliensis* suggest that the Mt replication and maintainability are constrained depend on the species and on the energy requirements regulated by the physiological and developmental status of organs.

Similarly, both genes of Nu and Mt genomes related to TCA cycle and oxidative phosphorylation have been predicted with high expression levels (Tables S10, S11). The results showed that 50, 58 & 54 enzymes coding genes of TCA pathway which are encoded by Nu genome, were found in *A. setaceus*, *A. taliensis* and *A. officinalis* respectively, in which 0, 3 & 5 of citrate synthases (CSs) (EC 2.3.3.1) genes, 4, 7 & 5 ATP citrate synthases (ACSs) (EC 2.3.3.8); and 7, 9 &10 malate dehydrogenases (MDHs) (EC 1.1.1.37) were found in *A. setaceus*, *A. taliensis* and *A. officinalis* respectively. Further analyses were conducted based on the phylogenic tree (Figure S5) of CSs and ACSs and their gene expression levels (Table S10). It was found that, except 2 pseudogenes of CSs (i.e., having trunked protein length with no expression in all sampled organs) in *A. officinalis*, there were 3 CSs & 4 ACSs, 3 CSs & 6 ACSs, and 0 CSs & 4 ACSs expressions in *A. officinalis*, *A. taliensis* and *A. setaceus* respectively. A total of 116, 122 & 120 enzyme genes was identified for oxidative phosphorylation pathway, in which 2 genes (i.e., *atp6* and *atp8*) coding for ATP synthase complex subunits were found from the Mt genome, while the other genes encoded by Nu genome were present in all 3 species. These coding enzymes were assembled into 5 super molecular complexes noted I to V (Table S11). The gene numbers of MDHs and CSs (Table S10) and the their expression levels were compared and the results showed that (Figure S6) the expression levels of MDHs and CSs genes in *A. officinalis* were higher than those in *A. taliensis* and *A. setaceus*, suggesting that the generation of citrate in *A. officinalis* was mainly based on OAA due to oxidation of S-malate, and the number of genes encoding the *A. officinalis* complex I for oxidative phosphorylation (with 35) were more than that in *A. taliensis* (with 32) and *A. setaceus* (with 31) (Table S11). The overall expression level of differentially expressed genes related to complex I encoded by *A. officinalis* was higher than that of *A. taliensis* and *A. setaceus*, but the number of genes encoding complex II encoded by *A. officinalis* (with 4) was less than that of *A. taliensis* (with 6) and *A. setaceus* (with 6), in which two of the four genes had extremely low expression levels (Fig. 5).

Discussion

After a long-term endophytic process, the Mt genome undergoes extremely complex changes in size and structure [51]. *A. officinalis*, *A. taliensis* and *A. setaceus* are not only representative species of dioecious species of sect. *Asparagus* and sect. *Archiasparagus* of the Subgenus of *Asparagus* and hermaphrodite species of subgen. *Asparagopsis*, but also represent species used as vegetable, herbal medicine as well as ornamental plants. Only *A. officinalis* had its whole Mt genome available, therefore the *A. taliensis* and *A. setaceus* species were selected to assemble and annotate the Mt genomes for both phylogenetic and functional analyses.

Base on the phylogenetic trees constructed with partial Mt genome encoding single copy genes (Fig. 1E) and whole Cp genomes (Fig. 1F), the species of sect. *Archiasparagus* lljin, which form a single clade had *A. taliensis* and *A. filicinus* grouped into a branch in the Mt tree, while Sect. *Asparagus* of the Cp tree grouped into a single clade, and *A. taliensis* and *A. filicinus* did not group into one clade. Bootstrap values of the Mt tree were relatively lower (<50%) than that of the Cp tree (bootstrap reaches 100%, Fig. 1), making Cp tree more reliable than the Mt tree, especially the position of *A. filicinus*, in accordance with the phylogenetic study of Bentz [52] on plastome sequences, that of Norup [53] on DNA barcodes as well as that of Leebens-Mack's group based on 1726 single nuclear loci for 318 accessions (representing 158 species of *Asparagus* L.) [54]. The inconsistency between the Cp and Mt trees in our study may be due to incomplete assembly of Mt genomes in certain species of *Asparagus* L., paternal leakage of mitochondrial genome [55], greater conservation of coding genes of Mt genome during evolution, as well as inconsistent selection of tree building genes during the construction of different organelle genome phylogenetic trees. Our Mt phylogeny indicates that using partial and whole Mt genomefor phylogenic analyses remains a challenge, and that developing a suitable protocol including using the coding, non coding genes as well as conserved intergenic sequences for phylogenetic tree construction may solve this problem.

HGTs in plant cells are commonly reported among Mt, Cp and Nu genomes [56, 57]. In the Nu genome of higher plants, sequences with similar or even identical fragments in Mt and Cp can often be found due to mutual HGTs [58]. The collinear analysis between Mt and Nu genomes showed that the HGTs between Mt and Nu of the three species were independent and reduced in dioecious species. The Nu and Mt genome of hermaphrodite plants may be more compatible due to selfing with less DNA variations in HGTs. However, we can hypothesize that the lower number of HGTs in dioecious species compared to the hermaphrodite species may result from a first phase of cross between dioecious species leading to more variations in genes of HGTs, followed by purification selection in the context of niche adaptation or domestication. Further KEGG enrichment analyses showed that the genes found in *A. officinalis* were related to organ development as well as signaling pathways, due to the fact that *A. officinalis* was domesticated thousands years ago as a vegetable (whose young tender stems are harvesting parts), having properties of quick biomass accumulation and higher yields of tender stems, these organ developments richen genes may have been selected to meet the requirement of quick energy and

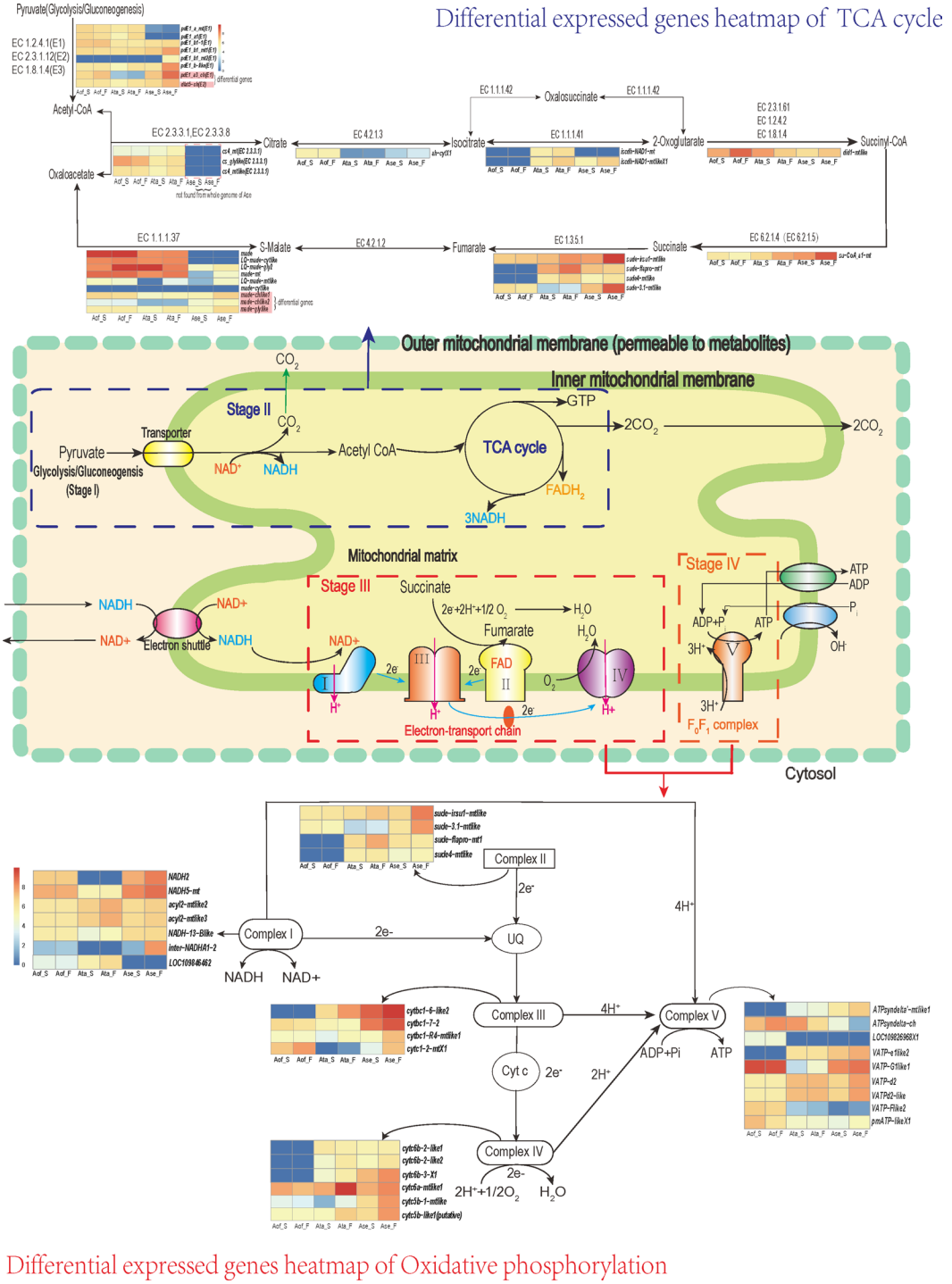

Fig. 5 DEGs heatmaps of TCA cycle and Oxidative phosphorylation; all genes are expressed differently in both stem and flower organs among the three species where *A. officinalis* have relatively higher expression levels in differentially expressed genes encoding EC1.1.1.37, EC2.3.3.1, EC2.3.3.8 and EC4.2.1.3 enzymes, while *A. setaceus* have relatively higher expression levels in differentially expressed genes encoding EC1.3.5.1 related enzymes

organic compound accumulations for rapid growth and development of harvested organs (young stems), while *A. taliensis* had genes related to saponin and flavonoid biosynthesis in accordance with the selection for medicinal purposes in *A. taliensis*. The UV damage repair like gene in *A. setaceus* for DNA repair may contribute to restoring DNA damage caused by occasional higher light with for restoring the DNA damages on certain level.

Based on the results of Mt and Nu collinear analyses (Figure S2), it is reasonable to speculate different

expression levels of homologous genes of both Mt and Nu genomes among the 3 *Asparagus* L species. The different expressions may be due to (i) the occurrence of incomplete HGTs genes among genomes, (ii) the redundant duplication of homologous genes in the Nu or Mt genomes accumulating more mutations and resulting in variable expression levels and (iii) the overall gene expression of both Nu and Mt concertedly contributing to maintain the function of energy production. For example *atp1* and *atp4* are homologous genes, but both Mt *atp4* and Nu genomes *atp4* (*Ata04G040320*) of *A. taliensis* were not expressed while the Mt *atp1* of *A. taliensis* had a higher expression level, leading to hypothesize that the *atp4* of both Mt and Nu genomic coding are pseudogenes and that the function of *atp4* in *A. taliensis* may almost be replaced by the function of *atp1* (Figure S2).

RNA editing is an indirect and highly specific repair mechanism of genetic variation which is extremely common in plant. The lack of RNA editing seriously affects the function of organelles and make plants unable to grow and develop well. The higher level of RNA editing in dioecious species than in *A. setaceus* may be due to the increase in DNA variation due to cross between dioecious plants, which is consistent with their higher copy numbers in Nu genes encoding RNA editing enzymes and their relatively higher expressions in dioecious species. While the RNA editing enzymatic system is encoded by the Nu genome, the Mt (maternally inherited) and Nu genome concertedly work on RNA editing in these species. The PPRs, MORFs, ORRMs, OZs and PPO1 are 5 enzymes for RNA editing, with PPRs having the highest gene number with higher expressions, while reduced gene numbers were detected in MORFs, ORRMs, OZs and PPO1s in all three species. This indicates that PPRs play a major role, while other families play minor (supplementary or complementary) roles in the Mt RNA editing [59, 60]. The changed RNA edition ratio, Nu coding RNA editing enzymes gene copies and their different expression levels may be partially related to the lower HGTs between Nu and MT detected in dioecious than in hermaphrodite species.

In membrane of Mt Complexes I to IV were used for oxidation in reducing the force of NADH and $FADH_2$, which were mainly derived from TCA reactions, while Complex V was a Mt type ATP synthase complex for ATP production powered by proton potentials. In plant, Complex I is a NADH dehydrogenase that oxidizes the NADH generated in the mitochondrial matrix, regenerating the oxidase form of NAD^+ to keep running TCA reactions. Complex II, is a succinate dehydrogenase component of the TCA, involved in the oxidation of succinate to fumarate. Both complex I and complex II transfer electrons to ubiquinone which is an abundant mobile electron transfer cofactor and used as shuttle for electrons transfer from complexes I and/or II to complex III. Complex III transfers electrons from ubiquinol to cytochrome C (cytC) which is the only protein in the electron transport chain (ETC) (Figure S6) to connect to complex IV (cytC oxidase), which is the terminal electron carrier in the ETC. The Complex V is an ATP synthase complex which use the proton potential to produce ATP. It was detected that the number of genes per complexes varied across species. We found that *A. officinalis* and *A. taliensis* accumulate more genes in complex I (NADH dehydrogenase complex), while *A. officinalis* has less genes in complex II (succinate dehydrogenase complex) than both *A. taliensis* and *A. setaceus.* This may indicate that they produce energy with different reducing forces (NADH and $FADH_2$).

Citrate is a key substrate driving the TCA cycle, mainly generated through two intermediate metabolites (acetyl-CoA, and oxaloacetic acid (OAA) by catalyzing CSs (EC 2.3.3.1). Acetyl-CoA is mainly produced through decarboxylation of pyruvate or β oxidation of fatty acid, while OAA is mainly produced by the oxidation of malic acid to oxaloacetic acid catalyzing by MDHs (EC 1.1.1.37) and, the processes of production of both acetyl-CoA, and OAA can generate NADH. Through comparative analysis of TCA cycle, and oxidative phosphorylation related genes, we didn't find CSs enzymes coding genes in *A. setaceus*, but only four ATP dependent citrate synthase enzymes (ACSs, Table S10), which means that the transition from acetyl-CoA to citrate in *A. setaceus* is mainly completed by ACSs in a non-efficient convertible reaction. However, the detection of higher expression levels and more copies of MDHs and CSs in *A. officinalis* suggests that this species can efficiently accumulate citric acid. Additionally, *A. officinalis* contains more genes with a higher expression of Complex I proteins, but less copies with a lower expression of Complex II. On the contrary, the gene copies and expression level of complex I is relatively low, but the gene copies and expression level of complex II is relatively high in *A. setaceus*. The results suggest that *A. setaceus* may mainly transmit electrons through the complex II with less efficacy for ATP production. However *A. officinalis* is mainly using complex I with higher efficiency for energy production, while *A. taliensis* appears to be in an intermediate state, using the balance of complex I and complex II to transfer electrons for ATP production. Therefore it can be speculated that *A. officinalis* mainly uses complex I for the electron transferring of oxidative phosphorylation, while *A. taliensis* has a high expression level and high number of gene copies in both complexes I and II, which may be due to the use of both balance of Complex I and Complex II to transfer electrons in this species. However, the number of Complex I genes in *A. setaceus* was the lowest among the 3 species, while its number of genes in complex II was

relatively high, indicating that the main way for electron transferring in *A. setaceus* respiratory chain may be conducted by complex II.

The energy released by the electrons transmitted by Complex I is more efficient than that released by Complex II under the same conditions, suggesting a decreasing energy efficiency (ATP production) from *A. officinalis* to *A. taliensis* and to *A. setaceus*. The different strategies and enzymatic systems for ATP production via TCA and oxidative phosphorylation among the three species would be related to higher and efficient energy requirements in *A. officinalis* as a vegetable with quick biomass accumulation in the context of ancient domestication, less biomass accumulation and shading tolerance in *A. setaceus* as a household ornamental plant, and secondary metabolites production in *A. taliensis* as a medicinal plant with recent domestication. As a vegetable crop, *A. officinalis* has been continuously cultivated and selected for a quick growth and development, what would have favoured the efficient Complex I pathway to obtain ATP in the context of high levels of energy required. The less efficient way to maintain the energy metabolism balance in *A. setaceus* is compatible with an ornamental plant which does not require intense energy consumption for rapid growth. The result suggests that domestication may affect the expression and variation of cytoplasmic genes, thereby altering the growth phenotype of plants in the same genus, as reported by Li et al. [61] who studied the variation of the chloroplast genome between wild and domestic ginseng.

Conclusion

The assembled and annotated circular Mt genomes of *A. taliensis* and *A. setaceus* provide a basis for further studying the evolution of energy production system of *Asparagus* L. in relation to domestication and to the adaptation to ecological niches.

Abbreviations

Mt	Mitochondrion
Mitogenome	Mitochondrial genome
TCA cycle	Tricarboxylic acid cycle
Cp	Chloroplast
Nu	Nuclear
HGTs	Horizontal gene transfers
PPRs	Pentatricopeptide-repeats
ONT	Oxford Nanopore Technologies
DGEs	Differential expression genes
IGV	Integrative Genomics Viewer
MORF	Multiple Organellar RNA Editing Factors
ORRM	Organelle RNA Recognition Motif
OZ	Organelle Zinc-finger
PPO1	Protoporphyrinogen IX Oxidase 1
MGICs	Mt genome integrated chromosomes
MAPK	Mitogen-activated protein kinase
MTQFs	MtDNA quality control factors
MTRFs	MtDNA replication factors
Cyt C	Cytochrome C
ETC	Electron transport chain
ACS	ATP citrate synthases
MDHs	Malate dehydrogenases
CS	Citrate synthases
OAA	Oxaloacetic acid
SNP	Single nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10768-3.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9
Supplementary Material 10
Supplementary Material 11
Supplementary Material 12
Supplementary Material 13
Supplementary Material 14

Acknowledgements

Not applicable.

Author contributions

ZCM and ZJL conceived and designed the research, HW and SEB contributed to writing and revising the manuscript. CL provided experimental materials for sequencing, SGW and WHDC performed experimental work. HW, YBL and ZJL performed the bioinformatics analysis. All authors have read and approved the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (32360089). Yunnan Agricultural joint Fund-key project (202101BD070001-027), Yunnan Basic Research Program (202305A0350012).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Collection of plant material

We have permission to collect *Asparagus* species. The collection of plant material complies with relevant institutional, national, and international guidelines and legislation.

Competing interests

The authors declare no competing interests.

Author details
[1]College of Agronomy and Biotechnology, Yunnan Agricultural University (YNAU), Kunming, Yunnan 650201, China
[2]Industrial Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu, Sichuan 610023, China
[3]Institute of Improvement and Utilization of Characteristic Resource Plants, YNAU, Kunming, China
[4]The Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province, Kunming, China

Received: 4 December 2023 / Accepted: 3 September 2024
Published online: 12 September 2024

References

1. Picault N, Hodges M, Palmieri L, Palmieri F. The growing family of mitochondrial carriers in Arabidopsis. Trends Plant Sci. 2004;9(3):138–46.
2. Feagin JE, Gardner MJ, Williamson DH, Wilson RJ. The putative mitochondrial genome of Plasmodium Falciparum. J Protozool. 1991;38(3):243–5.
3. Ward BL, Anderson RS, Bendich AJ. The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). Cell. 1981;25(3):793–803.
4. Sloan DB, Oxelman B, Rautenberg A, Taylor DR. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. BMC Evol Biol. 2009;9:260.
5. Lei Binbin L, Shuangshuang L, Guozheng W, Yumei S, Aiguo. Hua Jinping: evolutionary analysis of mitochondrial genomes in higher plants. Mol Plant Breed. 2012;10(04):490–500.
6. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15(8):e1008373.
7. Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, et al. Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol. 2004;136(3):3486–503.
8. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell. 2011;23(7):2499–513.
9. Wynn EL, Christensen AC. Repeats of unusual size in Plant mitochondrial genomes: identification, incidence and evolution. G3 (Bethesda). 2019;9(2):549–59.
10. Shi G, Cui Z, Hui M, Liu Y, Chan TY, Song C. Unusual sequence features and gene rearrangements of primitive crabs revealed by three complete mitochondrial genomes of Dromiacea. Comp Biochem Physiol Part D Genomics Proteom. 2016;20:65–73.
11. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. Mitochondrion. 2008;8(1):5–14.
12. Boonsom T, Waranuch N, Ingkaninan K, Denduangboripant J, Sukrong S. Molecular analysis of the genus Asparagus based on matK sequences and its application to identify A. racemosus, a medicinally phytoestrogenic species. Fitoterapia. 2012;83(5):947–53.
13. Seberg O, Petersen G, Davis JI, Pires JC, Stevenson DW, Chase MW, Fay MF, Devey DS, Jorgensen T, Sytsma KJ, et al. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. Am J Bot. 2012;99(5):875–89.
14. Saha PS, Ray S, Sengupta M, Jha S. Molecular phylogenetic studies based on rDNA ITS, cpDNA trnL intron sequence and cladode characteristics in nine Protasparagus taxa. Protoplasma. 2015;252(4):1121–34.
15. Moreno R, Castro P, Vrana J, Kubalakova M, Capal P, Garcia V, Gil J, Millan T, Dolezel J. Integration of genetic and cytogenetic maps and identification of sex chromosome in Garden Asparagus (Asparagus officinalis L). Front Plant Sci. 2018;9:1068.
16. Liu B, Li B, Zhou D, Wen X, Wang Y, Chen G, Li N. Steroidal saponins with cytotoxic effects from the rhizomes of Asparagus cochinchinensis. Bioorg Chem. 2021;115:105237.
17. Bopana N, Saxena S. Asparagus racemosus–ethnopharmacological evaluation and conservation needs. J Ethnopharmacol. 2007;110(1):1–15.
18. Jiang ZX, Geng L, Huang YD, Guan SA, Dong W, Ma ZY. The model of rough wetting for hydrophobic steel meshes that mimic Asparagus setaceus leaf. J Colloid Interface Sci. 2011;354(2):866–72.
19. Watharkar AD, Kadam SK, Khandare RV, Kolekar PD, Jeon BH, Jadhav JP, Govindwar SP. Asparagus densiflorus in a vertical subsurface flow phytoreactor for treatment of real textile effluent: a lab to land approach for in situ soil remediation. Ecotoxicol Environ Saf. 2018;161:70–7.
20. Kubota S, Konno I, Kanno A. Molecular phylogeny of the genus Asparagus (Asparagaceae) explains interspecific crossability between the garden asparagus (A. Officinalis) and other Asparagus species. Theor Appl Genet. 2012;124(2):345–54.
21. Liao RJ, Yang Y, Ye BH, Li N, Chen YW, Weng YF, Du GJ, Li HB. [Transcriptome analysis of rhizome of Polygonatum Cyrtonema and identification of candidate genes involved in biosynthetic pathway of steroidal saponin]. Zhongguo Zhong Yao Za Zhi. 2020;45(7):1648–56.
22. Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. Nat Commun. 2017;8(1):1279.
23. Li SF, Wang J, Dong R, Zhu HW, Lan LN, Zhang YL, Li N, Deng CL, Gao WJ. Chromosome-level genome assembly, annotation and evolutionary analysis of the ornamental plant Asparagus setaceus. Hortic Res. 2020;7(1):48.
24. Sheng W. The complete mitochondrial genome of Asparagus officinalis L. Mitochondrial DNA B Resour. 2020;5(3):2627–8.
25. Andrews S. FastQC A Quality Control tool for high throughput sequence data. http://www.bioinformaticsbabrahamacuk/projects/fastqc/ 2014.
26. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
27. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.
28. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
29. He W, Xiang K, Chen C, Wang J, Wu Z. Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. Brief Bioinform 2023, 24(1).
30. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
31. Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.
32. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963.
33. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
34. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. Methods Mol Biol. 2014;1079:131–46.
35. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
36. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
37. Wang Y, Li J, Paterson AH. MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. Bioinformatics. 2013;29(11):1458–60.
38. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative Toolkit developed for interactive analyses of big Biological Data. Mol Plant. 2020;13(8):1194–202.
39. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32(suppl2):W20–5.
40. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov (Camb). 2021;2(3):100141.
41. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinformatics. 2005;6:96.
42. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In.; 2013: arXiv:1303.3997.
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
44. Small I, Peeters N, Legeai F, Lurin C. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics. 2004;4(6):1581–90.
45. Rayapuram N, Hagenmuller J, Grienenberger JM, Bonnard G, Giege P. The three mitochondrial encoded CcmF proteins form a complex that interacts with CCMH and c-type apocytochromes in Arabidopsis. J Biol Chem. 2008;283(37):25200–8.

46. Yagi Y, Tachikawa M, Noguchi H, Satoh S, Obokata J, Nakamura T. Pentatricopeptide repeat proteins involved in plant organellar RNA editing. RNA Biol. 2013;10(9):1419–25.
47. Yang Y, Fan G, Zhao Y, Wen Q, Wu P, Meng Y, Shan W. Cytidine-to-uridine RNA editing factor NbMORF8 negatively regulates Plant Immunity to Phytophthora Pathogens. Plant Physiol. 2020;184(4):2182–98.
48. Shi X, Bentolila S, Hanson MR. Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in Arabidopsis. Plant Signal Behav. 2016;11(5):e1167299.
49. Gipson AB, Hanson MR, Bentolila S. The RanBP2 zinc finger domains of chloroplast RNA editing factor OZ1 are required for protein-protein interactions and conversion of C to U. Plant J. 2022;109(1):215–26.
50. Zhang F, Tang W, Hedtke B, Zhong L, Liu L, Peng L, Lu C, Grimm B, Lin R. Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. Proc Natl Acad Sci U S A. 2014;111(5):2023–8.
51. Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci U S A. 2015;112(33):10177–84.
52. Bentz PC, Liu Z, Yang JB, Zhang L, Burrows S, Burrows J, Kanno A, Mao Z, Leebens-Mack J. Young evolutionary origins of dioecy in the genus Asparagus. Am J Bot. 2024;111(2):e16276.
53. Norup MF, Petersen G, Burrows S, Bouchenak-Khelladi Y, Leebens-Mack J, Pires JC, Linder HP, Seberg O. Evolution of Asparagus L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. *Mol Phylogenet Evol* 2015, 92:25–44.
54. Bentz PC, Burrows JE, Burrows SM, Mizrachi E, Liu Z, Yang J-B, Mao Z, Popecki M, Seberg O, Petersen G et al. Bursts of rapid diversification, dispersals out of southern Africa, and two origins of dioecy punctuate the evolution of Asparagus. *bioRxiv* 2024:2024.07.25.605174.
55. Wang N, Li C, Kuang L, Wu X, Xie K, Zhu A, Xu Q, Larkin RM, Zhou Y, Deng X, et al. Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. Proc Natl Acad Sci U S A. 2022;119(43):e2206076119.
56. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A. 2002;99(15):9905–12.
57. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5(2):123–35.
58. Niu Y, Gao C, Liu J. Complete mitochondrial genomes of three Mangifera species, their genomic structure and gene transfer from chloroplast genomes. BMC Genomics. 2022;23(1):147.
59. Sun T, Shi X, Friso G, Van Wijk K, Bentolila S, Hanson MR. A zinc finger motif-containing protein is essential for chloroplast RNA editing. PLoS Genet. 2015;11(3):e1005028.
60. Takenaka M, Zehrmann A, Verbitskiy D, Kugelmann M, Hartel B, Brennicke A. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc Natl Acad Sci U S A. 2012;109(13):5104–9.
61. Li MR, Shi FX, Li YL, Jiang P, Jiao L, Liu B, Li LF. Genome-wide variation patterns uncover the origin and selection in cultivated ginseng (Panax ginseng Meyer). Genome Biol Evol. 2017;9(9):2159–69.

Publisher's note