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

He Wu
Yunnan Agricultural University (YNAU)
Wenhua Dongchen
Yunnan Agricultural University (YNAU)
Yunbin Li
Yunnan Agricultural University (YNAU)
Sylvia E Brown
Yunnan Agricultural University (YNAU)
Shugu Wei
Sichuan Academy of Agricultural Sciences
Chun Lin
Yunnan Agricultural University (YNAU)
Zichao Mao (zmao@ynau.edu.cn)
Yunnan Agricultural University (YNAU)
Zhengjie Liu
Yunnan Agricultural University (YNAU)

Research Article

Keywords: Asparagus L., mitochondrial genome, phylogenetic analysis, RNA editing, energy production system

Posted Date: December 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3703463/v1

Additional Declarations: No competing interests reported.

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) genome (mitogenome) remains largely unknown. In this study, the typical herbel medicine of *A. taliensis* and ornamental plant of *A. setaceus* were used to assemble and annote the mitogenomes, and the resulted mitogenomes were further compared with published mitogenome of *A. officinalis* for the analyses of their functions due to 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 coding proteins 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 transferrings (HGTs) between Nu and Mt genomes among these species. Based on RNAseq data, the Mt RNA editing were detected and *atp1* and *ccmB* RNA editing of *A. taliensi*s were futher confirmed by DNA sequencing. Simultaneously homologous search found 5 Nu coding gene families including pentatricopeptide-repeats (PPRs) involved in Mt RNA editing were predicted in these species. Finally, the Mt genome variations, gene expressions and mutural HGTs between Nu and Mt were detected with correlation to their growth and developmental phenotypes respectively. The results suggest that, both Mt and Nu genomes coevoluted to maintain the Mt organella replication and meet requirements of energy production through TCA and oxidative phosphorylation among these species.

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

Highlights

1.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*.

2.Independent horizontal gene transfers (HGTs) between the mitochondrial and nuclear genomes were detected with reduce in dioicous species, while Mt RNA editing rate is higher in dioecious species with consistency of their higher copies and higher expression levels of involved nuclear coding RNA editing enzyme gene families.

3.Detected different gene copies and expression levels of both pathways of TCA and oxidative phosphorylation genes indicates the different efficiency of citric acid accumulatation and ATP synthesis due to adaptation and or domestication among species of *A. officinalis*, *A. taliensis* and *A. setaceus.*

Background

Mitochondria, as energy factories, exists in almost all eukaryotic cells providing energies to power cells by continuously producing adenosine triphosphate (ATP) using both tricarboxylic acid cycle (TCA cycle) and oxidative phosphorylation pathways to maintain the energy requirements of all biological activities[1]. With the development of DNA sequencing and bioinformatic technology, 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 multicurricular 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, normally the long reads sequence data needing ONT or PacBio platforms. Additionally due to adaptative 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 origination center, and southwestern China, regarded as a center of diversity for dioecious species. There are 8 dioecious species including, *A. taliensis* which is 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* [16] and *A. racemosus* [17] are traditional medicinal plants used in China and India respectively. *A. setaceus* and *A. densiflorus* are hermophrodite species originating from Africa and are used as ornamental plants cultivated and radiated 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 dioeicous 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 a herbal medicinal plant (by natives of Yunnan) having higher content of steroidal saponins accumulated in its roots system, while *A. setaceus* 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 evolution analysis 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 construct the species phylogenetic relationship of *Asparagus L*, ii) collinearly analyze the Mt and Nu genomes inferring the mutual HGTs between Nu and Mt genomes among the 3 species, iii) compare annotated Mt genomes and gene expressions with transcriptome data to predict and reverify possible RNA editing sites and the main candidates involved in Nu coding gene families for Mt RNA editing respectively and iv) detect the Mt genome variations, differential expression genes (DGEs) and changed HGTs, correlating to metabolism, phenotypes for analyzing the co-evolution of Nu and Mt genomes, maintaining the energy requirements for adaptation and domestication in their environmental niches respectively.

Materials and Methods

Plant material and sequencing

The green variety “Guelph Millennium” of *A. officinalis* (Aof_G) and the wild type male of *A. taliensis* (Ata_M), were obtained from 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 replications and store at -80˚C, or used directly for DNA and RNA extraction for DNA and RNA sequencing, respectively.

Approximately 5–10 μg of total DNA from each sample was used to build a shotgun library with an average insertion size of 150 bp. An Illumina Genome Analyzer II (Illumina, USA) was used to sequence the DNA samples in paired-end sequencing mode. The generated clean data were 4 ~ 30 GB for each sample, and for third-generation sequencing, the library protocol for Nanopore PromethION (Nanopore, Oxford) sequencing was used to obtain ~ 75 GB clean data. All DNA sequencing data have been uploaded to the China National Center for Bioinformation (CNCB) with ID CRA007986 (https://ngdc.cncb.ac.cn/gsa/s/39tkmg2u),CRA008000 (https://ngdc.cncb.ac.cn/gsa/s/k5b77Vxs) and CRA009175 (https://www.cncb.ac.cn/search?dbId=&q=CRA009175).

RNAseq and data processing

RNA was extracted from the roots, stems, and flowering twigs of Aof_G and Ata_M for transcriptome sequencing. At the same time, both Ilumina and third-generation sequencing (with projects ID: PRJNA564485) and transcriptome data (SRR10177391, SRR10186988 and SRR10187001) of *A. setaceus* were downloaded from NCBI and all raw data were evaluated with FastQC [25] then filtered with Fastp [26] to get clean data with default parameters. The clean reads were further mapped to the corresponding species reference Nu and Mt genone using hisat2 v2.2.1 [27]. Expression quantification was performed through [28] was used 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

CSAT [29]were used for Mt genome 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, if the third generation DNA sequencing data are not available. Racon v1.4.21[31] was used to self-correct the assembled contig, using Nanopore sequencing data, the entire process is iterated three times, and pilon v1.24 [32] was used to align the Illumina data to the corrected sequence, the entire process is iterated twice. Finally, genome function annotation and mapping were performed using the Geseq online website (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html#), with the reference genome of Mt genome of *A. officinalis* and all parameters of the software setting as default.

Analyses of phylogeny

Based on the assembled or downloaded Mt and Cp genomes of the genus *Asparagus L*., 19 Mt single copy gene coding protein 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 analyzed by OrthoFinder [33] to obtained pseudo protein sequence of each species respectively to perform multiple sequence alignments using mafft [34]. After, trimming the gap using trimAl [35], the Mt phylogenetic tree was finally constructed by the maximum likelihood method 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/).

Colinearity analysis of Mt genome among species

Based on the General Feature Format (GFF), annotation file and protein sequences, the Mt genomes of each species were directly analyzed for collinearity using MCScanX [37] with default parameters respectively, followed by the visualized results of collinearity 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 (https://cran.r-project.org/web/packages/pheatmap) was used to visilized the genes expression patterns.

Prediction and verification of RNA editing sites

The obtained coding gene sequence of Mt genome was submit to PREP-mt online prediction software [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* where selected for separate amplification for both DNA and cDNA to verify the prediction results of RNA editing using PCR primer listed in Table S6.

RNA editing enzymes gene family analysis

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 (hmmsearch 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 candiates 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].While RAxML was used to construct a phylogenetic tree with 1000 bootstrap replicates. The protein sequences under the same branch as the reference gene were obtained to motif analysis using MEME Suite v5.5.1 (https://meme-suite.org/meme), while the localization of signal prediction was performed using Predotar v.1.0.4 [44] (https://urgi.versailles.inra.fr/predotar/) on all proteins. The eggNOG was used to annotate the gene function, and the expression heatmaps was visalized using pheatmap package of RStudio.

Genes related to Mt function and their expression analysis

All genes related to Mt biosynthesis, 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 *Arabidopis thalianta*, *Oryza sativa* and *A. officinalis* were obtained with seqkit (https://github.com/shenwei356/seqkit) respectively for database of BLAST analyses. The Nu and Mt genomes coding proteins of *A. officinalis*, *A. taliensis* and *A. setaceus* were used as query to screen candidate homoloudous genes respectively with critical seclection standards of identity rate > 0.4, coverage rate > 0.8. The gene list related to Mt biosynthesis, oxidative phosphorylation and citrate cycle of *A. officinalis*, *A. taliensis* and *A. setaceus* were obtained for comparison and gene expression analysis 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 S1), having 3 ~ 4-fold size of their assembled complete circular Cp genomes respectively (Fig. 1B

and D). The assembled Mt genomes of *A. taliensis* and *A. setaceus* were annotated with 53 and 52 genes respectively which are similar with reported 53 genes of Mt of *A. officinalis* (Table S2, 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 respectively only being assembled from Illumina sequencing data using GetOrganella as described in materials and methods (Figure. 1E, Table S1). The GC contents of the resulted assembled Mt genomes of *Asparagus L*. species ranged between 46.58%~46.62%. Choosing the hermaphrodite species *A. setaceus* as outgroup, the phylogenetic trees were constructed using the conserved 19 coding proteins 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 respectively. The Mt phylogenetic tree results showed that, the major dioecious species *of Asparagus L* were group into 2 major clades in which one major clade consisting of 4 species (i.e. *A. dauricus, A. longifiorus, A. persicus A. oligoclonos*) is sister with the clade consisting of all other dioecious species, in which, *A. officinalis* and *A. taliensis* are grouped into two independent smaller clades (Fig. 1E). The Cp phylogenetic trees also showed that all dioecious species are grouped into 2 major clades in which *A. lycopodineus*, *A. meioclados*, *A. taliensis* and *A. cochinchinensis* are group into one major clade, which is sister to the other left dioecious species, in which *A. filicinus* is an independent clade, which is sister to the clade consisting of the other dioecious species including *A. officinalis* (Fig. 1F). According to the traditional taxonomy, *A. setaceus* belonged to *Asparagus* subgenus. *Asparagopsis* (labeled as green) is a hermaphrodite species, while dioeicous species are classified *into* 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 sect. *Archiasparagus* lljin were grouped into a single clade, which is a 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 are not group into a single clade as compared to that sect. *Asparagus* which had all sampled species of grouped into a single clade. It is interesting to note that species of *A. filicinus* which belongs to sect. *Archiasparagus* lljin is sister with all species of sect. *Asparagus* as an independent smaller clade. The resulted Mt phylogenetic trees also showed species within sect. *Archiasparagus* lljin, grouped into a branch which is sister the clade consisting of *A. meioclados* and *A. lycopodineus.* While in the phylogenetic tree of Cp genomes, *A. taliensis* and *A. filicinus* are not merged into one clade. Based on the Mt phylogenetic tree, species of *A. officinalis*, *A. angulofractus* and *A. neglectus* are grouped as a clade which is sister to clades of sect. *Archiasparagus* lljin, while based on Cp phylogenetic tree, only *A. officinalis* and *A. neglectus* are group into a smaller clade which is a sister clade to all other sampled species in the sect. *Asparagus.* The currently constructed phylogenetic trees of both Cp and Mt genomes, *A. setaceus*, *A. taliensis* and *A. officinalis* which are representing the hermaphrodite and dioecious species of the sect. *Archiasparagus* and sect. *Asparagus* respectively, simultaneously these 3 species are also representative species used as 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 are 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) repectively, while ATMT and AOMT have 15 collinear blocks. Further analyses of these block containing genes related to enegry production by TCA and oxidative phosporylation in AOMT showed that *Aof-cox2, Aof-ccmFc, Aof-atp8, Aof-cox3 and Aof-nad1* are specific beweeen AOMT and ATMT, while genes: *Aof-cox1, Aof-atp6, Aof-nad9, Aof-ccmC* are specific beween AOMT and ASMT. Comparing ATMT with AOMT and ASMT respectively, the genes: *Ata-atp1, Ata-ccmFn, Ata-ccmB* are specifc beweeen ATMT and AOMT, while genes: *Ata-nad4, Ata-nad6, Ata-cox1*, *Ata-cox3* and *Ata-atp4* are specific beween ATMT and ASMT. Similarly, the genes for direct enegry production: *Ase-atp1*, *Ase-atp9*, *Ase-atp8*, *Ase-ccmFn*, *Ase-ccmB* are specific beetween ASMT and AOMT, while *Ase-cox2*, *Ase-cox3*, *Ase-nad4* are specific between 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 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 homolgous genes in both ATMT (i.e. *Ata-ccmC*, *Ata-ccmB* and *Ata-ccmFn*) and AOMT (*Aof-ccmC*, *Aof-ccmFn* and *Aof-ccmB*) are detected in the three species(Table S2). The results also showed inversions or arragements of gene clusters between Mt genomes among species, for example, ATMT and ASMT have 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 mitgenomes of the 3 species may be due 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 using MCScanX software. 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* respectively (Figure S2 A, B and C). The Nu genomes among the 3 species also have colinear regions of chromosomes in which Chr01 ~ Chr10

of A. officinalis are more homologous with *Chr01 ~ Chr10* of *A. taliensis*, while the fragments of Mt genome integrated chromosomes (MGICs) (i.e. AseChr01 and AseChr03) are 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*) having the integrated gene *rps7* is the same as the *rps7* in *AseChr01*(*A. setaceus*), while all HGTs between Nu and Mt genomes among the 3 species are 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 the different energy requirement for certain growing environmental niches respectively. These differently detected collinear blocks containing genes in Nu genomes of each species including 5 genes each flanked to its 5' (up) and 3'(down) blocks respectively were combined for further KEGG richening analyses. The richening analyses results showed that these genes are 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), other genes corresponding to multiple specific metabolic or signaling pathways were also found. It is needed to note that 4 enriched genes including 2 *CYP450s* (*Ata04G040340* and *Ata05G023230*) and 2 transcriptional factors (*Ata05G011840* and *Ata04G040190*) of *A. taliensis* were annotated to be involved in biosynthesis of secondary metabolites such as: isoflavonoids and steroids. The results of these enriching genes in *A. taliensis* that positively correlated with the species was chosen and recently cultivated as medicinal plant which have accumulated more secondary metabolites such as: steroidal saponins and isoflavonoids (Fig. 2, Figure S2 and Table S3). It is also interesting to note that, with a total of 24 genes in *A. officinalis*, 1/3 of *them* (i.e: *AofChr08.983*, *AofChr08.970*, *AofChr08.981*, *AofChr03.1816, AofChr03.1814*, *AofChr03.1819*, *AofChr09.700* and *AofChr09.702*) are annotated to relate to the metabolic and signal transduction pathway in which 6 genes are involved in plant growth and development of plant organs (e.g. Mitogen-activated protein kinase (MAPK) as well as signaling pathway). Due to the fact that *A. officinalis* was chosen and domesticated for thousands of years 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 or domesticated to meet the requirement of quick energy and organic compound accumulations for rapid growth and development of harvest organs (young stems). 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 for DNA repair which may contribute to its shading tolerance proprieties for being domesticated as household ornamental plants with relatively slow growth and low light (especially UV) environmental condition.

After comparing the expression levels of both Mt genes and their corresponding homologous Nu genes among species (Figure S2), it was found that some Nu genome genes for Mt function have relatively low expression levels, while the corresponding homologous genes in Mts exhibit relatively high expressions, and the average overall expression level tends to stabilize among species. Based on these results, (even though molecular mechanisms are not known), it is reasonable to speculate that, the different expression levels of homologous genes between Nu and Mt genome may be due to i) incomplete genes of HGTs between genomes, ii) the redundant duplication of homologous genes in the Nu or Mt genomes accumulating more mutations during evolution and results in changing their expression levels and iii) the overall gene expressions 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* have been detected without expression, while the Mt *atp1* of *A. taliensis* had a higher expression level inferring that the *atp4* of both Mt and Nu genomic coding are pseudogenes or the function of *atp4* in *A. taliensis* is almost replaced by the function of its *atp1* (Figure S2).

RNA editing of Mt genes and their involved enzymes coding gene families of Nu genome

The RNA editing sites of Mt RNA were predicted with online software PREP-Mt (http://www.prep-mt.net/). The results showed that, 36 out of 39 coding Mt genes in AOMT and ATMT were found to have 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 were found to have RNA editing with average SPG of 15.1 (Table S4). 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 respectively (Table S4). The detected RNA editing type results showed that, 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 without detecting the third Nt of the codon (Table S4). 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 S4 & Table S5). To reverify the detected results of RNA editing sites, the *Ata_atp1* and *Ata_ccmB* of *A. taliensis were selected* for PCR amplification with both Mt genomic DNA and derived cDNA as templates by designed primers listed in Table S6 followed by DNA sequencing. The DNA sequencing results showed, CDS of *Ata_atp1* in 1168, 1178 &1262 sites and CDS of *Ata_ccmB* in 313, 338, 392, 406, 424 & 428 sites were reverified 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 with critical standards of 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 RNA recognition motifs (ORRMs) [48], organella zinc-fingers (OZs) [49] and protopor phyrinogen 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*, 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 S7). 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* have 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 neofunctionalization compared to *A. officinalis* and *A. setaceus*. The high gene number of PPRs and reduced number of MORFs, ORRMs, OZs and PPO1s in all three species indicates that PPRs take major roles, while other families take minor (with supplementary or complementary) roles in Mt RNA editing. Futher 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, Ase03G183, 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. Higher RNA edition rates in both dioecies species (*A. taliensis* and *A. officinalis*) than the hermaphrodite species *A. setaceus*, may be due to easy accumulation of the genomic variation through *hybridation* of males and females for offsprings production, which is consistent with their higher copy numbers in Nu genome encoding RNA editing enzymes and relatively higher expressions in dioecious species than species of *A .setaceus* which accumulates less genomic DNA variation due to the production of offsprings through selfing.

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 organella, TCA and oxidative phosphorylation in *Arabidopsis thaliana*, *Oryza sativa* ssp. *japonica and A. officinalis*, were used as database to homlogously search the genes of the 3 species of *Asparagus L* through BLAST as described in materials and methods. The results indicate that, the gene numbers related to Mt biosynthesis and assembling in *A. setaceus* (with total 196*)* are less than both counterparts of *A. officinalis* (207) and *A. taliensis* (258) (Tables S8) in which ASMT, AOMT and ATMT encoded 35, 36 and 36 proteins which had expressions at least in 1 sampled organ respectively (Table S8). The major expansion or contraction of gene or gene families 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*, while 84 MTQFs and 77 MTRFs in *A. setaceus* respectively. There are more expansions of MTRFs genes in dioecious species (*A. officinalis* and *A.taliensis*) than the couterparts of hermaphrodite species suggesting higher copy numbers of Mt organelles in cells of dioecius species than hermophodite species. It is interesting to note that the *atp1* and *atp9* for conding subunit of ATP synthase complex in Mt of *A. setaceus* having duplicated copies in its Nu genome respectively, while, *atp1* and *atp4* (subunit of ATP synthase complex) have duplication partial gene fragment(*atp1*) and completement copies(*atp4*) in Nu genome of *A. taliensis* (Figure S2 and Table S8) however, no ATP synthase subunit coding genes had been detected in *A. officinalis*, additionally, as described, *atp4* of *A. taliensis* were detected with no expression in both Mt and Nu genomes, indicating that, the demostication or adaptative evolution forced the progressive removal of duplication copies of ATP synthase complex enzyme genes either in Nu or Mt genome of *Asparagus L* for efficient ATP production through oxidative phosphorylation. Futher anlyses of gene expressions of MTQFs and MTRFs based on RNAseq data showed that a total of 180 genes have 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) are common in all organs, while, 13, 9 & 20 DEGs are special in Rs, Ss and Fs respectively (Figure S3) among them, 21 differentially expressed genes are upregulated in *A. officinalis* (Figure S4), from the expression level it can be seen that the overall expression level of AOMT replication factor genes is 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 decided by the requirements of energies 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 among species with high expression levels (Tables S9, S10). The results showed that 50, 58 & 54 enzymes coding genes of TCA pathway which are encoded by Nu genome, are found in *A. setaceus, A. taliensis and A. officinalis, in which* 5, 3 & 0 of citrate synthases (CSs) (EC 2.3.3.1) genes, 5, 7 & 4 ATP citrate synthases (ACSs) (EC 2.3.3.8); and 10, 9 & 7 malate dehydrogenases (MDHs) (EC 1.1.1.37) were found in *A. officinalis*, *A. taliensis* and *A. setaceus* respectively. Further analyses were conducted based on the constructed phylogenic tree (Figure S5) of CSs and ACSs and their gene expression levels (Table S9), it was found that, except 2 pseudogenes of CSs, (judged by having trunked protein length with no expression in all sampled organs) in *A. officinalis*, there are 3 CSs & 4 ACSs, 3 CSs &6 ACSs, and 0 CSs & 4 ACSs expressions in *A. officinalis*, *A. taliensis* and *A. setaceus* respectively For oxidative phosphorylation pathway, a total of 116,122 & 120 enzyme genes was identified, in which 2 genes (i.e. *atp6*

and *atp8*) coding 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 of I to V (Table S10) in which complexes of I to IV were used for oxidation of reducing force of NADH and $FADH_2$ which were mainly derived from TCA reactions, while Complex V were Mt type ATP synthase complex for ATP production powered by proton potentials. In plant, Complex I is an NADH dehydrogenase that oxidizes the NADH generated in the mitochondrial matrix, regenerating the oxidate form of NADH (NAD^+) to keep running TCA reactions. Complex II, which is succinate dehydrogenase as a component of enzymes of the TCA, by 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 shuttles 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) (Fig. 5) 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 genes among five complexes have variation among species where the *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* which may indicate their energy production with different reducing force (NADH and $FADH_2$) which are derived from different pathways among species. This result is consistent with dioecious species had more CSs genes, while hermaphrodite species *A. setaceus* only containing ACSs which are mainly used for splitting the citrate into OAA and Acetyl-CoA. In the TCA cycle, citrate is a key substrate driving the TCA cycle, mainly generated through two intermediate metabolites (acetyl-CoA, and oxaloacetic acid (OAA) by catalyzing of 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. The gene numbers of MDHs and CSs (Table S9) and the their expression levels were compared and the results showed that (Fig. 5), 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 the 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 S10). The overall expression level of differentially expressed genes related to complex I encoded by *A. officinalis* is higher than that of *A. taliensis* and *A. setaceus* (Fig. 6), but the number of genes encoding complex II encoded by *A. officinalis* (with 4) is less than that of *A. taliensis* (with 6) and *A. setaceus* (with 6) in which two of the four genes have extremely low expression levels. It can be speculated that *A. officinalis* mainly uses complex I for the electron transfer of oxidative phosphorylation, while in *A. taliensis*, there are related high expression level and gene copies of 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* is the least among the 3 species, while the number of genes in complex II is relatively high, which may be the main way for electron transfer in *A. setaceus* respiratory chain 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. As a vegetable crop, *A. officinalis* has been continuously cultivated and domesticated, the evolution force to choose the Complex I pathway to obtain the efficient ATP to fill the large energy gap required for its quick growth and development. As an ornamental plant, *A. setaceus* does not require intense energy consumption for rapid growth, so it may have chosen a more gentle or less efficient way to maintain its energy metabolism balance. *A. taliensis* has been recently cultivated and domesticated, its domestication history is much shorter than *A. officinalis*, therefore, it not only retains the metabolic characteristics of the original species *A. setaceus*, but also evolves its complex I genes as *A. officinalis* There is a balanced way by using both complex I and complex II pathways for its own respiratory chain and energy production.

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 Subgenus of Asparagus and hermaphrodite species of subgen.*Asparagopsis* but also represent crops used as ornamental, herbal medicine as well as vegetable purposes. Only *A. officinalis* has its whole Mt genome available, the *A. taliensis* and *A. setaceus* species were selected to assemble and annotate the Mt genomes for both phylogeny and functional analyses.

Base on the phylogenetic tree constructed by 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 has *A. taliensis* and *A. filicinus* grouped into a branch in the Mt phylogenetic tree, while Sect. *Asparagus* of the Cp phylogenetic tree was grouped into a single clade, with *A. taliensis* and *A. filicinus* not merged into one clade. According to the bootstrap values, that of the Mt tree is relatively low (< 50%) than that of the Cp tree (bootstrap reaches 100%), making the classification results of the Cp tree is more reliable than Mt tree on the position of *A. filicinus.* This inconsistency between the Cp and Mt trees may be due to incomplete assembly of Mt genomes in certain species or more variations in the Mt genes occurring during evolution (Table S1, Fig. 1).

HGTs in plant cells are commonly reported among Mt, Cp and Nu genomes [52, 53]. 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 [54]. The collinear analysis between Mt and Nu genomes

shows that the HGTs between Mt and Nu of the three species were independent and reduced in dioecious species. KEGG enriching analysis showed that the genes found in *A. officinalis* were related to organ development as well as signaling pathways while *A. taliensis* had genes related to saponin and flavonoid synthesis indicating the evolution and domestication to select faster growth for the vegetable crop *A. officinalis* and medicinal using of *A. taliensis*. The UV damage repair like gene in *A. setaceus* for DNA repair may contribute to its shading tolerance proprieties and UV sensitivity for it being domesticated as household ornamental plants

RNA editing is an indirect and highly specific repair mechanism of genetic variation which is extremely common in plant, the lack of RNA editing will seriously affect the function of organelles and make plants unable to grow and develop well. The more RNA editing in dioecious species than *A. setaceus* may be due to easy accumulation of DNA variation due to hybridization between different dioecious plants. The hermaphrodite *A. setaceus* have relatively less accumulated DNA variation may be due to the generation of offspring through self-pollination. Mitochondria are inherited maternally, while the RNA editing enzymatic system are encoded by the Nu genome, the Mt and Nu genome concertedly work on RNA editing in these species. The PPRs, MORFs, ORRMs, OZs and PPO1 are 5 enzymes for RNA editing,withPPRs having the highest gene number with higher expressions, while reduced numbers are detected in MORFs, ORRMs, OZs and PPO1s in all three species indicates that PPRs take major roles, while other families take minor (with supplementary or complementary) roles in the Mt RNA editing.

Through comparative analysis of MTRFs, TCA cycle, and oxidative phosphorylation related genes, the expression of MTRFs genes in *A. officinalis* were found to be relatively high among the three species, at the same time, *A. officinalis* was found to have fewer multi-copy genes (5) among mitochondrial replication related genes, while *A. taliensis* and *A. setaceus* had a large number of redundant genes (29&13 respectively) (Table S8). This may be a strong domestication and selection made by human domestication to remove almost all redundant genes to reduce the burden on the entire Nu and Mt genome with the highest expression of gene either by Nu or Mt single copy genes. For genes related to TCA and oxidative phosphorylation, CSs enzymes and their coding genes were not found in *A. setaceus*, only four ATP dependent citrate synthase enzymes (ACSs, Table S9), which means that the transition from acetyl-CoA to citrate in *A. setaceus* is mainly completed by ACSs with its nonefficient convertible reaction. However, the detection higher expression levels and more copies of MDHs and CSs in *A. officinalis* suggest *A. officinalis* can efficiently accumulate citric acid. Additionally, *A. officinalis* contains more genes and higher expression of Complex I proteins but less copies and 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. setaceu*s may mainly transmits 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, in which both complex I and II are selected through the balance of complex I and complex II to transfer electrons for ATP production. The different strategies and enzymatic systems for ATP production via TCA and oxidative phosphorylation among the three species are due to, higher and efficient energy requirement in *A. officinalis* as a vegetable for quick biomass accumulation through strong and long historical domestication, less biomass accumulation and shading tolerance in *A. setaceus* as a houshold ornamental plants and medicinal plant with time short domestication in *A. taliensis.*

Conclusion

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

List of abbreviations

Abbreviation	Full name
Mt	Mitochondrion
Mitogenome	Mitochondrial genome
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

Declarations

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.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the China National Center for Bioinformation repository with accession numbers: C_AA045705.1 - C_AA046744.1, CRA007986, CRA008000 and CRA009175 (https://www.cncb.ac.cn/)

Funding

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

Competing interests

The authors declare that they have no competing interests

Authors' Contributions

Zichao Mao and Zhengjie Liu conceived and designed the research, He Wu and Sylvia E Brown contributed to writing and revising the manuscript. Chun Lin provided experimental materials for sequencing, Shugu Wei and Wenhua Dongchen performed experimental work. He Wu, Yunbin Li and Zhengjie Liu performed the bioinformatics analysis. All authors have read and approved the manuscript.

Acknowledgements

Not applicable

References

1. Picault N, Hodges M, Palmieri L, Palmieri F: **The growing family of mitochondrial carriers in Arabidopsis**. *Trends Plant Sci* 2004, **9**(3):138-146.
2. Feagin JE, Gardner MJ, Williamson DH, Wilson RJ: **The putative mitochondrial genome of Plasmodium falciparum**. *J Protozool* 1991, **38**(3):243-245.
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, Li Shuangshuang, Liu Guozheng, Wang Yumei, Su Aiguo, Hua jinping: **Evolutionary Analysis of Mitochondrial Genomes in Higher Plants**. *Molecular Plant Breeding* 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):3486-3503.
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):2499-2513.
9. Wynn EL, Christensen AC: **Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution**. *G3 (Bethesda)* 2019, **9**(2):549-559.
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 Proteomics* 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):947-953.
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):875-889.
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):1121-1134.
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):866-872.
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,

161:70-77.

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):345-354.
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):1648-1656.
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):2627-2628.
25. Andrews S: **FastQC A Quality Control tool for High Throughput Sequence Data**. *http://wwwbioinformaticsbabrahamacuk/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-i890.
27. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nat Methods* 2015, **12**(4):357-360.
28. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.
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):737-746.
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**:131-146.
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):1972-1973.
36. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**(9):1312-1313.
37. Wang Y, Li J, Paterson AH: **MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans**. *Bioinformatics* 2013, **29**(11):1458-1460.
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):1194-1202.
39. McGinnis S, Madden TL: **BLAST: at the core of a powerful and diverse set of sequence analysis tools**. *Nucleic Acids Research* 2004, **32**(suppl_2):W20-W25.
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**. *Innovation (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):1792-1797.
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):1581-1590.
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):25200-25208.
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):1419-1425.
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):2182-2198.

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):215-226.
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):2023-2028.
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):10177-10184.
52. 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):9905-9912.
53. Timmis JN, Ayliffe MA, Huang CY, Martin W: **Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes**. *Nat Rev Genet* 2004, **5**(2):123-135.
54. 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.

Figures

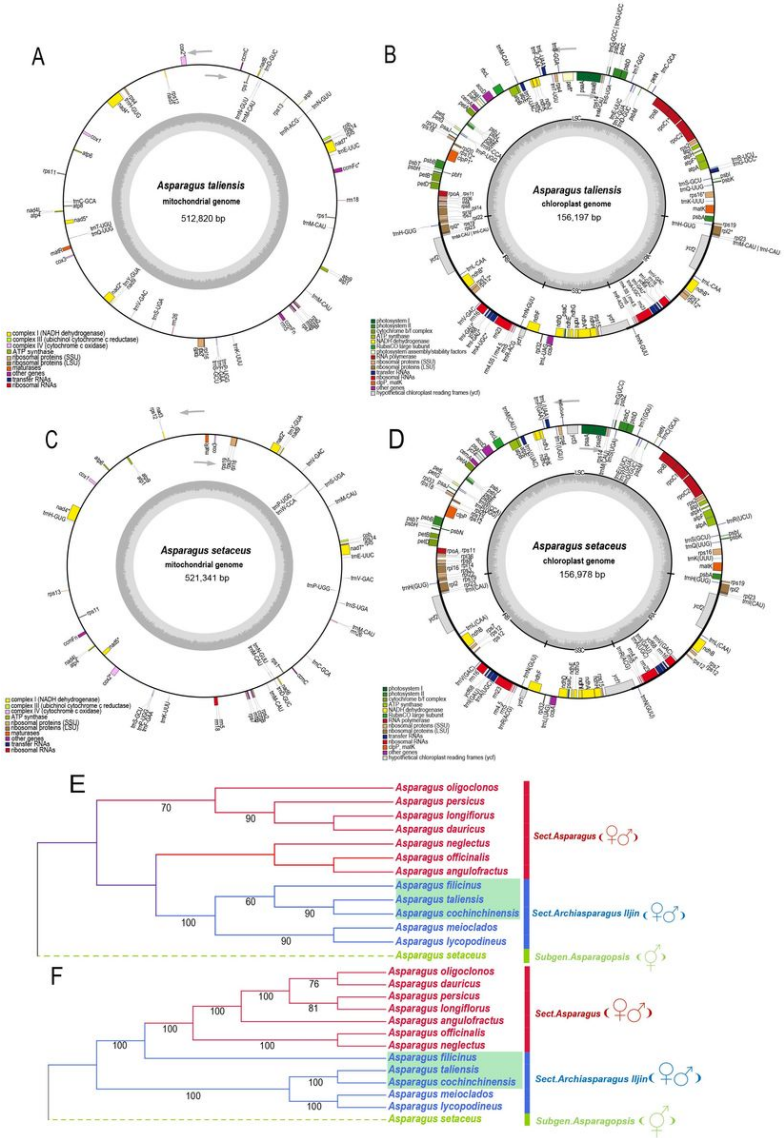

Figure 1

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.

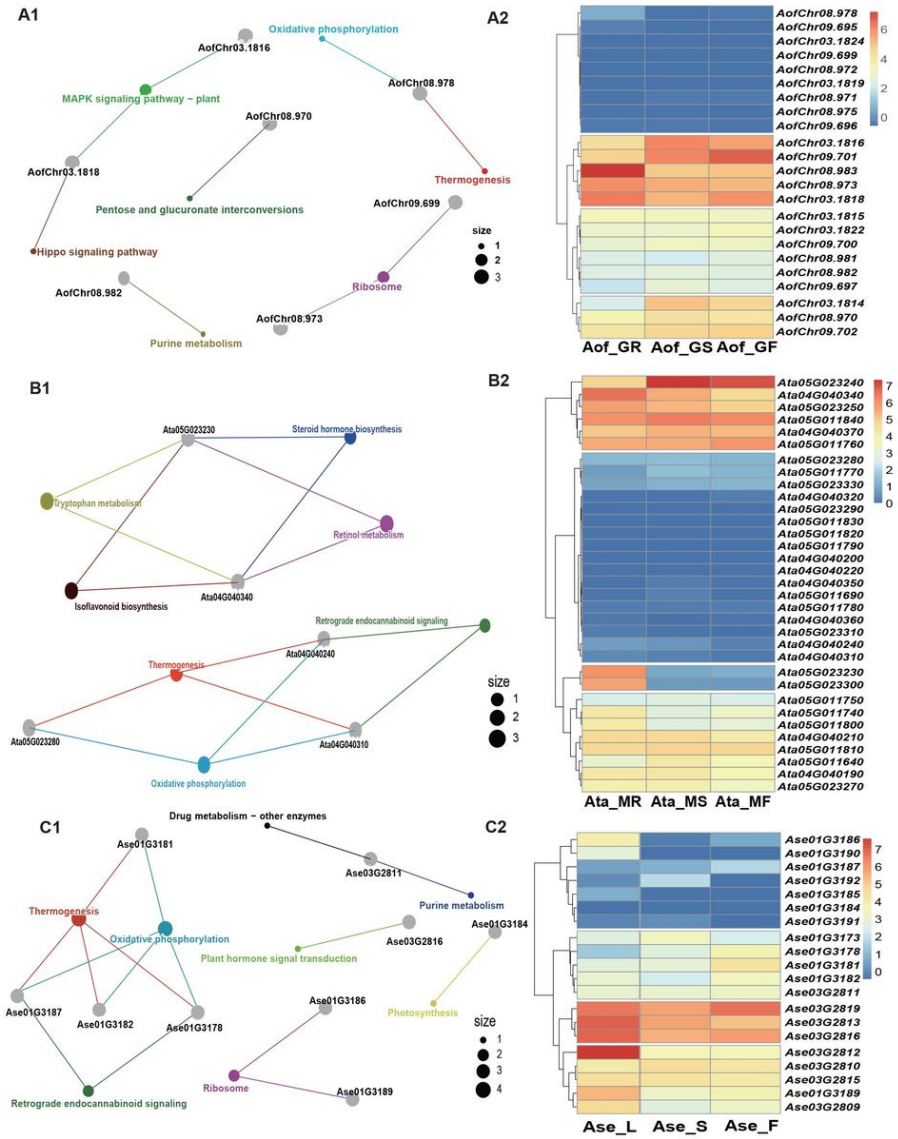

Figure 2

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 down flanking genes respectively; **A2-C2**: collinear Nu genes heatmap expressions of *A. officinalis, A. taliensis and A. setaceus* with their five up and down flanking 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*

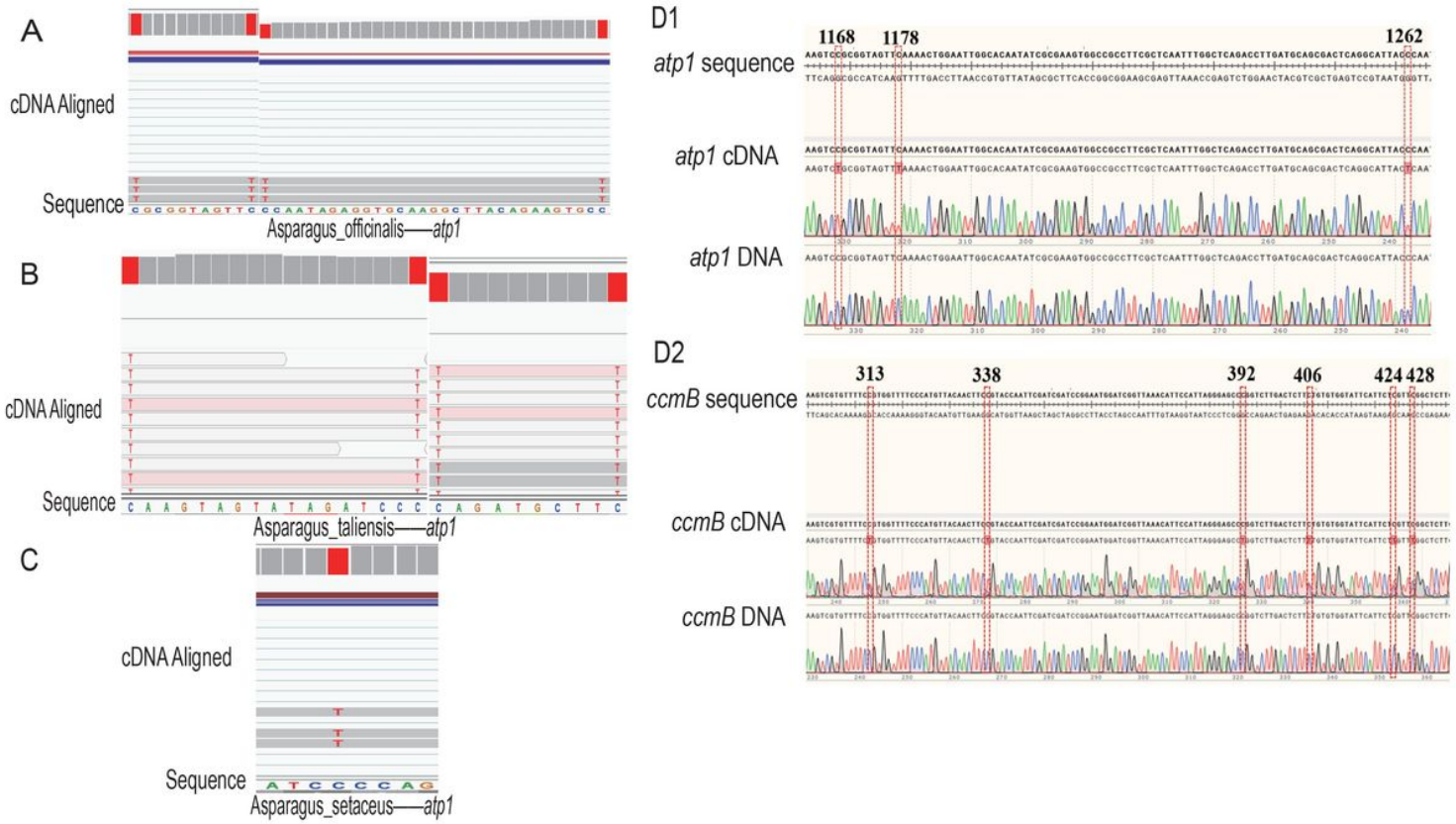

Figure 3

RNA editing prediction and validation; **A-C**: RNA editing prediction of *A. officinalis*, *A. taliensis* and *A. setaceus atp1* gene, the black borderindicates 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.

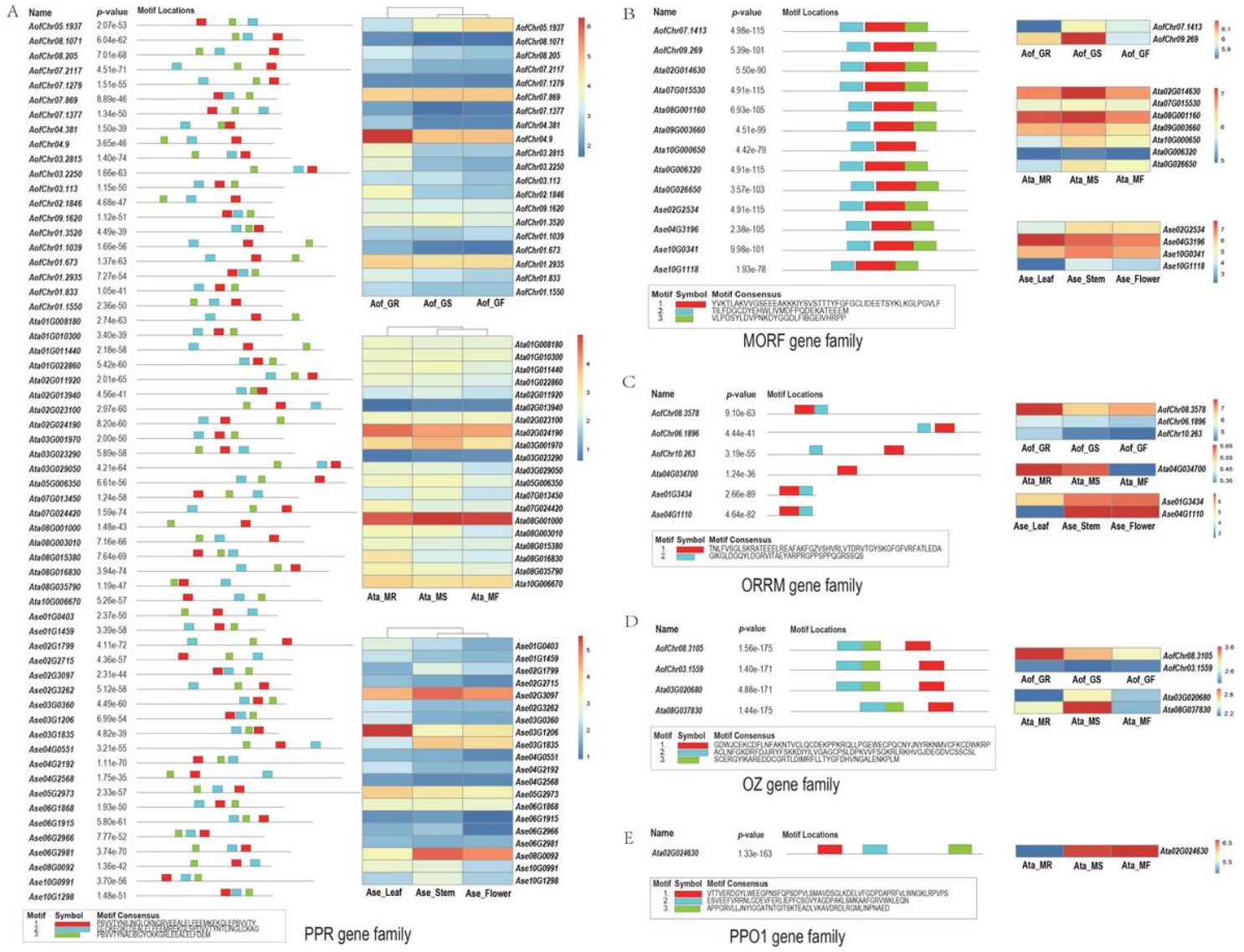

Figure 4

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).

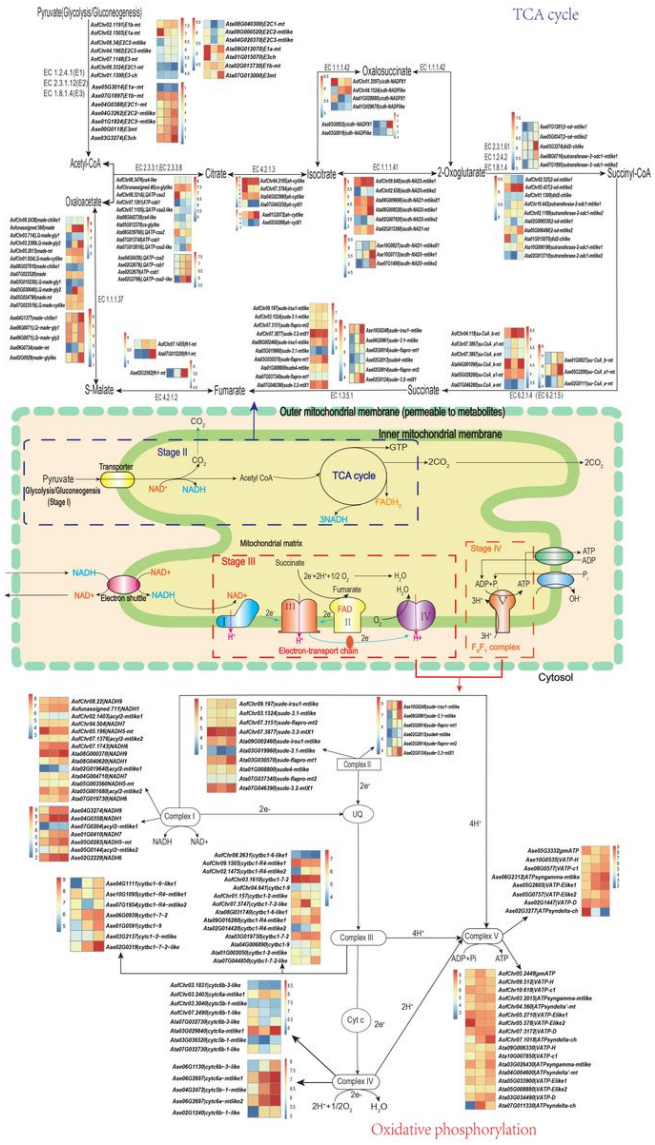

Figure 5

Energy metabolism pathways related to mitochondrial function; the TCA cycle process (upper) is shown in a dark blue box, and oxidative phosphorylation (lower) is shown in red and orange boxes in the Mt model diagram(middle), I - IV represent the complex I to complex IV in electron transport chain, V represent the ATP synthase. Each step is connected by a unidirectional (represent irreversible reaction) arrow or bidirectional (represent reversible reaction) arrow in TCA cycle and oxidative phosphorylation pathway, and the number on the arrow represents the Enzyme Commission (EC) number of the enzyme catalyzed to this metabolism reaction. For the functions and detailed expression levels of all genes refer to the attached tables TS9 and TS10. The heatmaps of Asparagus officinalis (Aof) and Asparagus taliensis (Ata) show the expression levels in roots, stems, and flowers from left to right, while the heatmaps of Asparagus setaceus (Ase) show the expression levels of leaves, stems, and flowers from left to right, all expression data are the average of three biological replicates for each organ.

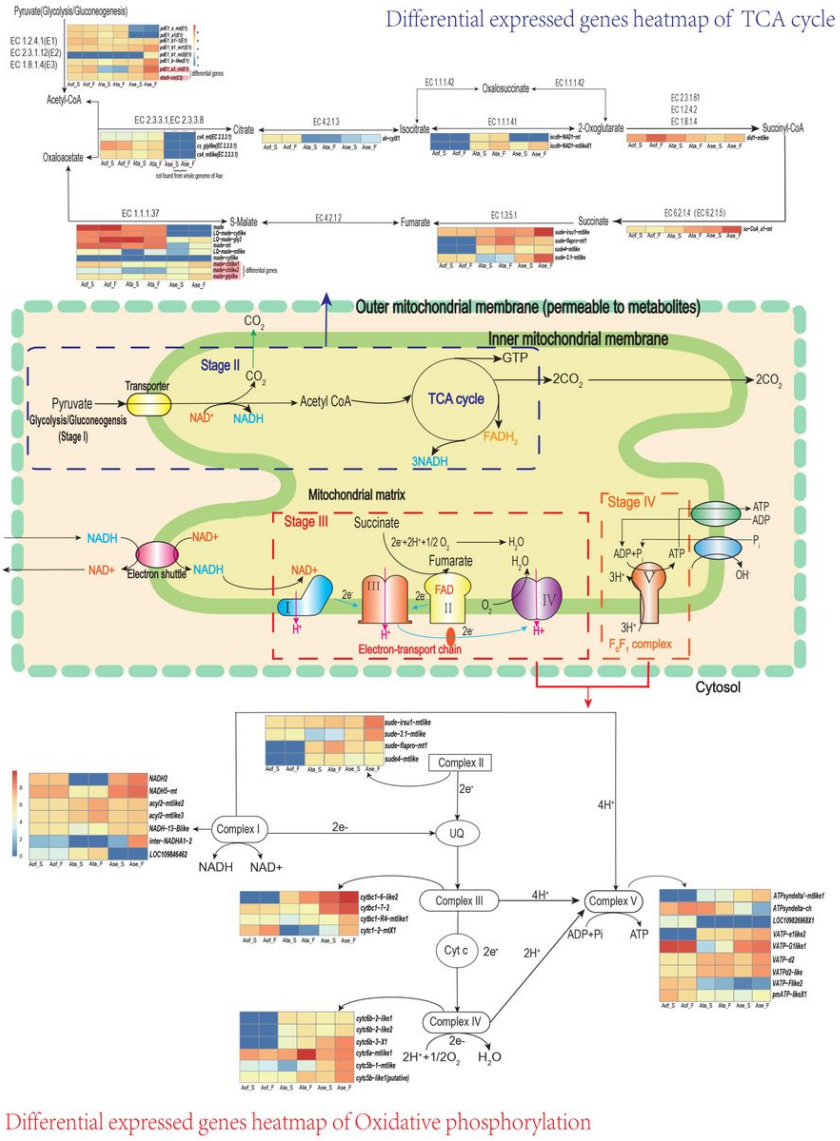

Figure 6

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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1Colinearloopdiagramofmitochondrialgenomeamongspecies.pdf
- FigureS2.CollinearityandexpressionofMtandNugenomes.pdf
- FigureS3.VenndiagramofTCAcyclereplicationfactorandoxidativephosphorylationdifferentialgenesinA.officinalisandA.taliensis.pdf
- FigureS4.HeatmapofTCAcyclereplicationfactorandOxidativephosphorylationfordifferentialgenesinA.officinalisA.taliensisandA.setaceus.pdf
- FigureS5.PhylogenetictreeofcitratesynthaseandATPcitratesynthasegeneamongA.officinalisA.taliensisandA.setaceus.pdf
- TableS1256.docx
- TableS3.NugenesofCollinearmtDNAandfivenucleargenesofupflankinganddownflankinginAsparagusLandtheirexpressionlevels.xlsx
- TableS4.PredictedRNAeditingsiteofgenusAsparagus.xlsx
- TableS7.MtRNAeditingrelatedtogenefamilygenesTPM.xlsx

- TableS8.MtDNAreplicationandqualitycontrolfactors11.12.xlsx
- TableS9.AllgenesandexpressionlevelsintheTCAcycle.xlsx
- TableS10.AllOxidativephosphorylationnucleargenesandtheirexpressionlevels.xlsx