

RESEARCH ARTICLE

Open Access



Unprecedented organelle genomic variations in morning glories reveal independent evolutionary scenarios of parasitic plants and the diversification of plant mitochondrial complexes

Yanxiang Lin¹, Pan Li², Yuchan Zhang³, Delara Akhter^{3,4}, Ronghui Pan^{3,5}, Zhixi Fu⁶, Mingqing Huang¹, Xiaobo Li^{7,8} and Yanlei Feng^{7,8*} 

Abstract

Background: The morning glories (Convolvulaceae) are distributed worldwide and produce economically important crops, medicinal herbs, and ornamentals. Members of this family are diverse in morphological characteristics and trophic modes, including the leafless parasitic *Cuscuta* (dodders). Organelle genomes were generally used for studying plant phylogeny and genomic variations. Notably, plastomes in parasitic plants always show non-canonical features, such as reduced size and accelerated rates. However, few organelle genomes of this group have been sequenced, hindering our understanding of their evolution, and dodder mitogenome in particular.

Results: We assembled 22 new mitogenomes and 12 new plastomes in Convolvulaceae. Alongside previously known ones, we totally analyzed organelle genomes of 23 species in the family. Our sampling includes 16 leafy autotrophic species and 7 leafless parasitic dodders, covering 8 of the 12 tribes. Both the plastid and mitochondrial genomes of these plants have encountered variations that were rarely observed in other angiosperms. All of the plastomes possessed atypical IR boundaries. Besides the gene and IR losses in dodders, some leafy species also showed gene and intron losses, duplications, structural variations, and insertions of foreign DNAs. The phylogeny reconstructed by plastid protein coding sequences confirmed the previous relationship of the tribes. However, the monophyly of 'Merremieae' and the sister group of *Cuscuta* remained uncertain. The mitogenome was significantly inflated in *Cuscuta japonica*, which has exceeded over 800 kb and integrated massive DNAs from other species. In other dodders, mitogenomes were maintained in small size, revealing divergent evolutionary strategies. Mutations unique to plants were detected in the mitochondrial gene *ccmFc*, which has broken into three fragments through gene fission and splicing shift. The unusual changes likely initially happened to the common ancestor of the family and were caused by a foreign insertion from rosids followed by double-strand breaks and imprecise DNA repairs. The coding regions of *ccmFc* expanded at both sides after the fission, which may have altered the protein structure.

*Correspondence: fengyanlei@outlook.com

⁸ Institute of Biology, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang, China

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



Conclusions: Our family-scale analyses uncovered unusual scenarios for both organelle genomes in Convolvulaceae, especially in parasitic plants. The data provided valuable genetic resources for studying the evolution of Convolvulaceae and plant parasitism.

Keywords: Convolvulaceae, *Cuscuta*, Plastid genome, Mitochondrial genome, *ccmFc*, Horizontal gene transfer

Background

Convolvulaceae Juss., also known as the morning glories and bindweeds, is a large family belonging to the order Solanales of the eudicots. It contains approximately 1900 species from 59 genera [1]. Members of the family include crops (e.g., sweet potato), vegetables (e.g., water spinach), medicinal plants (e.g., dried seeds of dodders and vines of *Erycibe*), ornamentals (e.g., morning glory, cypress vine, and moonflower), and tough weeds (e.g., dodders). Convolvulaceae is cosmopolitan and exhibits a rich diversity of morphological characteristics comprising herbs, shrubs, trees, climbers, and leafless parasitic plants *Cuscuta* (dodders).

In the past decades, species in Convolvulaceae were classified into 12 tribes and six clades using several DNA markers [2–5]. With great advances in genome sequencing technologies, plastid genomes (plastomes) of many plant taxa were completed. However, the plastome studies of Convolvulaceae mainly focused on *Ipomoea* [6–8] and dodders [9–12]. Many genera of this family were neglected, leaving the phylogenetic relationship untested with larger datasets and the positions of the ‘Merremieae’ and Cuscutaeae unresolved [4, 13]. Additionally, only two mitochondrial genomes (mitogenomes) in this family were known, one is recently published *Evolvulus alsinoides* [14] and the other is *Ipomoea nil*, which was obtained from the whole genome sequencing project [15].

On the other side, organelle genomes in Convolvulaceae exhibited multiple peculiar characteristics. For example, the intron deletion of *rpl2* in Convolvulaceae plastomes was a synapomorphy and a unique event in Asteridae [3]. The *infA* has been lost from the plastome, and inverted repeat (IR) boundaries have been altered in *Ipomoea* and dodders [8, 12]. The mitogenome of *I. nil* is only 266 kb in length, which is much smaller than most known mitogenomes in angiosperms. Moreover, annotation of the mitogenome (GenBank ID: AP017303) reveals a very unusual *ccmFc* gene, with fragments located in three different positions. Further studies are required to reconcile these findings. Plastomes of dodders, like many other heterotrophic plants (such as [16, 17]), exhibited smaller size, reduced genes, variable structure, and elevated nucleotide substitutions [10, 11, 18]. The nuclear genomes of dodder also experienced massive gene losses [19, 20]. In contrast,

the evolutionary pattern of their mitogenomes remains unexplored.

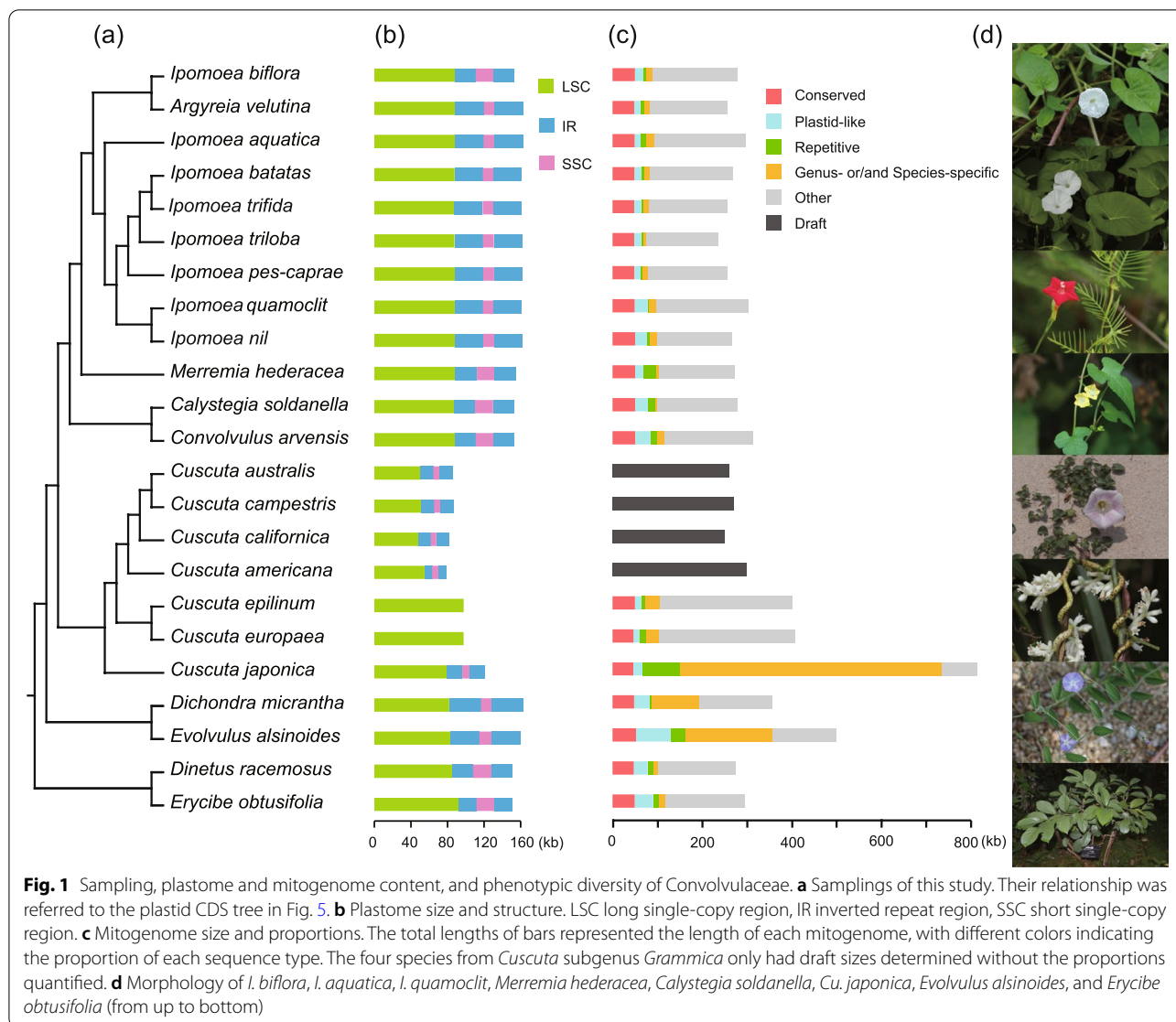
To unlock the secrets of organelle genomes of Convolvulaceae, we assembled 22 mitogenomes and 16 plastomes. Our collections covered five of the six clades and eight of 12 tribes, representing one of the most comprehensive samplings of this group. We unveiled the unusual organelle genome variations and depicted their evolutionary trajectories. Mitogenomes in parasitic dodders have evolved towards different directions, either acquiring large amounts of foreign DNA or retaining a small size. Unusual mutations were detected in the mitochondrial *ccmFc* in all family members, and omics data were employed to confirm the gene fission, splicing shift, and coding expansion.

Results and discussion

Plastome variations in Convolvulaceae

Our assemblies include at least 12 new plastomes, among which six genera were sequenced for the first time (Additional file 1). We also assembled four dodders for the first time, two from the subgenus *Cuscuta* (*Cu. epilinum* and *Cu. europaea*), and two from the subgenus *Grammica* (*Cu. americana* and *Cu. californica*). These new plastomes showed similar features to the known ones in *Cuscuta*, including the losses of genes, introns, and IR regions (Fig. 1b; Additional file 1) [12, 21]. In the leafy species, the plastid DNA sequences or the gene content did not differ much, and the main distinction came from the variation of the IR regions (Figs. 1b and 2a; Additional file 1). IR regions expanded or shrank in different directions and resulted in different architectures—all the species in this family were atypical in IR boundaries (Fig. 2a). *Dinetus* and *Dichondra* contained some inversions, and they were the only two leafy species that lost the gene *rpl23* (Fig. 2b; Additional file 1). The *Dichondra* plastome also had some unusual duplications, which could create another copy of *rpl16* (Fig. 2b, c, blue and red ribbon represented the original and second copy of *rpl16*, respectively). *Rps16* and *rpoC1* in *Evolvulus*, and *rps16* and *ycf3* in *Dichondra* have experienced intron losses. *InfA* was lost in Convolvulaceae, and BLASTP searches in Convolvulaceae nuclear peptides have confirmed the transfer to nucleus, like many other species [22].

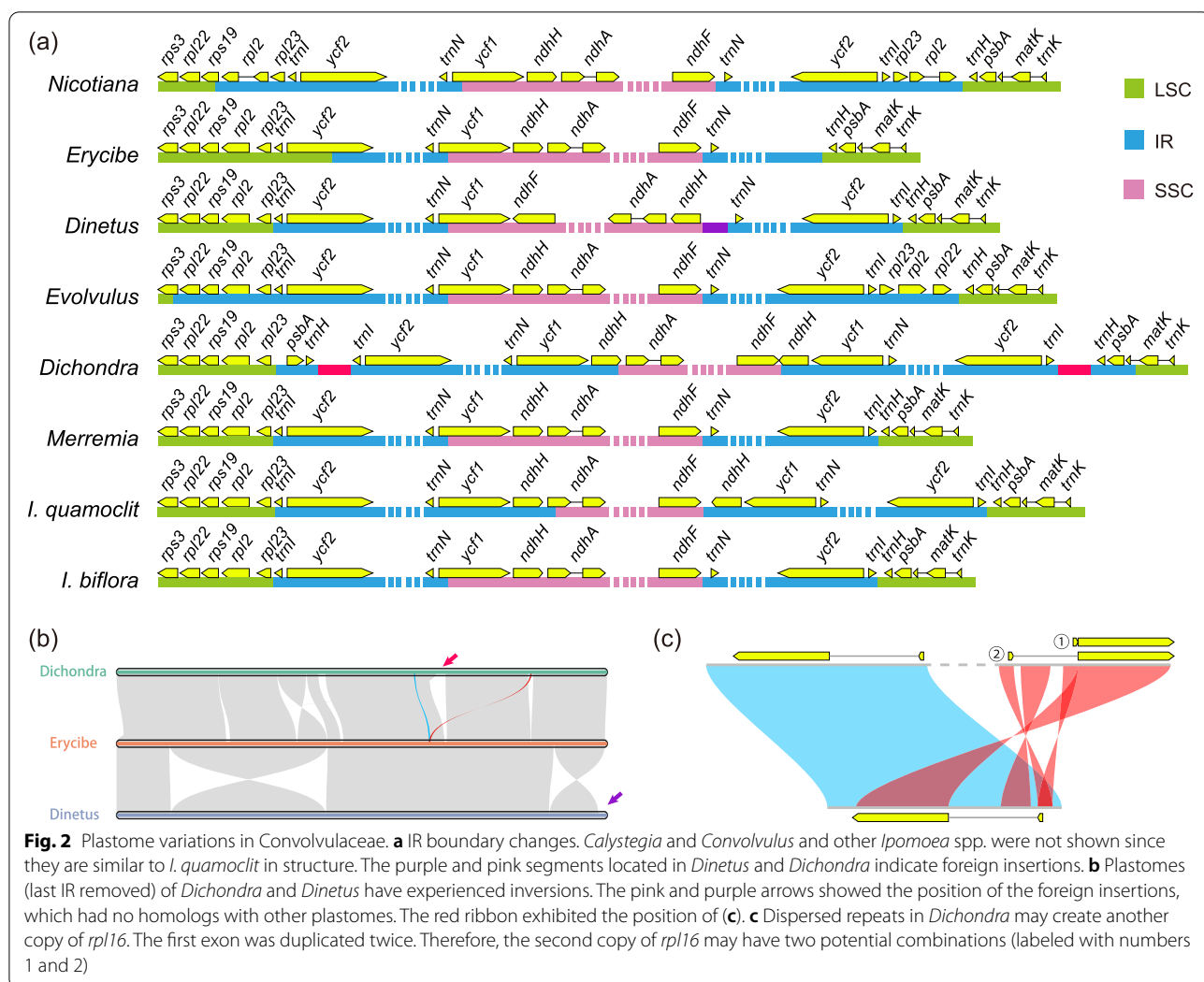
Dinetus and *Dichondra* had a *ca.* 1.5 and 2.2 kb foreign DNA inserted into the short single-copy



(SSC) region and the IR region, respectively (Fig. 2a, b, purple and red bars and arrows). These two foreign sequences had no homologs in Convolvulaceae mitogenomes, nor were they similar to any sequences in NCBI *nt* database (no significant hits were yielded from BLASTN searches), so their origins were mysterious. Foreign DNAs in plastomes were mainly known to be transferred from mitogenome, such as in Apiaceae [23–25], Apocynaceae [26], Anacardiaceae [27], Orobanchaceae [28], and Poaceae [29–32], while transfers from other sources were rare [25]. The foreign insertions in *Dinetus* and *Dichondra* plastomes could represent two additional examples of non-mitochondrial origin.

Mitogenome variations in Convolvulaceae

We assembled mitogenomes of 22 species in the family Convolvulaceae, making this one of the few studies comprising a large number of new and complete mitogenomes in plants (Fig. 1c, Additional file 2). Complete mitogenomes were obtained for 18 of these, including all leafy species, *Cuscuta epilinum*, *Cu. europaea*, and *Cu. japonica*, while only drafts were obtained for the four dodders from subgenus *Grammica* (*Cu. americana*, *Cu. australis*, *Cu. californica*, and *Cu. campestris*) due to large amounts of repeats. Mitogenomes in Convolvulaceae are mostly around 300 kb in length, which is shorter than most angiosperms, including their close relatives Solanales. The gene content of Convolvulaceae mitogenomes is similar to other angiosperms



(Additional file 2). The “core” genes are well preserved except *ccmFc* (see below). Compared to Solanales, *rpl2* and *sdh3* were missing in all family members investigated. The *rps7* was only found in *I. biflora*, and phylogenetic analysis indicated that it was closest to Rosales (bootstrap value 99% with *Morus notabilis* and *Cannabis sativa*; Additional file 3, *rps7*). Therefore, the *rps7* most likely have been lost in the common ancestor of Convolvulaceae and *I. biflora* re-gained it by horizontal gene/DNA transfer (HGT). Mitogenomes of the *Grammica* were also the most degenerated in the family—they lost more genes, similar to their plastomes. *Cu. epilinum*, *Cu. europaea*, and *Cu. japonica* showed no significant differences in gene content (Additional files 1 and 2). Most mitochondrial genes in dodders had increased substitution rates, but none was from HGT (Additional file 3).

In contrast to the generally small size in the family, the mitogenome in *Cu. japonica* has exceeded 800

kb, reaching twice as large as others (Fig. 1c). Besides, *Dichondra*, *Evolvulus*, *Cu. epilinum*, and *Cu. europaea* also showed expansions. Pair-wise mitogenomic synteny of the tribes was compared, which revealed rapid changes in mitogenomic structure and DNA content (Fig. 3). To understand the causes of the difference, we dissected the mitogenomic sequences into five classes: conserved (have homologs in all Convolvulaceae), plastid-like (potential plastid insertions), repetitive, genus-, or/and species-specific and other (the rest) (Fig. 1c). Only a small proportion of DNA exists in all (mostly intragenic regions; Fig. 1c, the “Conserved” proportion). The most significant difference was from genus- or/and species-specific sequences (GSS, details see “Methods” section), especially in *Cu. japonica*.

To identify the potential origin, GSS of *Dichondra*, *Evolvulus*, and *Cu. japonica* were blasted against *nt* database and screened the best hits (Additional file 4). The best hits were further grouped into orders. Although

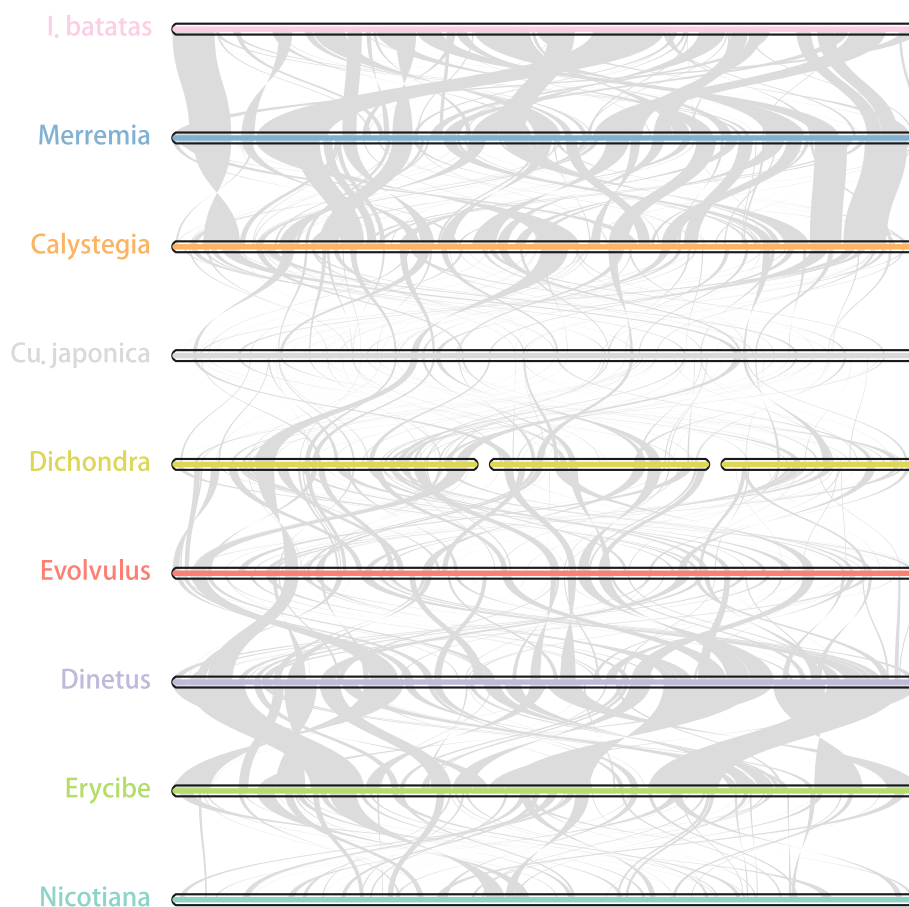


Fig. 3 Pairwise mitogenome synteny among tribes. The bars indicated the mitogenomes, and the ribbons showed the homologous sequences between the adjacent species. *Dichondra* has three chromosomes. Only one species of each tribe was used, and their order referred to the plastid CDS tree in Fig. 5. Tobacco (*Nicotiana tabacum*, GenBank: NC_006581) was also compared

many of them had unknown origins (i.e., had no results from the BLAST searches), most GSS shared high similarity with distantly related taxa (Fig. 4a). Fabales, Solanales, Lamiales, Malpighiales, and Gentianales occupied large proportions in the GSS in all the three species, which may indicate that their expansion has experienced similar events. In contrast, Caryophyllales and Santalales in *Evolvulus* and Rosales, Sapindales, Apiales, etc., in *Cu. japonica* could imply independent evolution.

Divergent evolution of the mitogenomes in dodders

The parasitic plants have evolved at least 12 times independently in angiosperms [33, 34], which provides an excellent model to study plant interactions and molecular evolution. Their genomes often show characteristics not commonly observed in other plants, including gene losses and HGTs [19, 20, 35–37]. Plastomes in green autotrophic plants are very conserved, while those in parasitic plants always show reduced size, loss of photosynthesis genes, different IR boundaries, and ascending

substitution rates [38], as reported for *Cuscuta* [10, 11, 18]. For mitogenomes, many obligate parasitic plants exhibited extensive HGTs from hosts, even including the replacements of the native genes, such as in *Cynomorium* (Cynomoriaceae [39, 40];), *Lophophytum mirabile* (Balanophoraceae [41, 42];), *Viscum album* (Santalaceae [43];), and *Aeginetia indica* (Orobanchaceae [44];). Therefore, it might not be surprising that *Cu. japonica* received large amounts of DNA from other species. However, mitogenome size in other dodders still resembled that of leafy species. An interesting question is how large the common ancestor of dodders was in mitogenome size—if the mitogenome of *Grammica* dodders degraded or *Cu. japonica* expanded (Fig. 4b, c). We hypothesized that if the mitogenomes shrank from big to small, there should contain some remnants, even though the remnants were very short. We used the significant HGT-like sequences of *Cu. japonica* (Additional file 4, length >500 bp) as queries to search for homologs in the other six dodders and then calculated

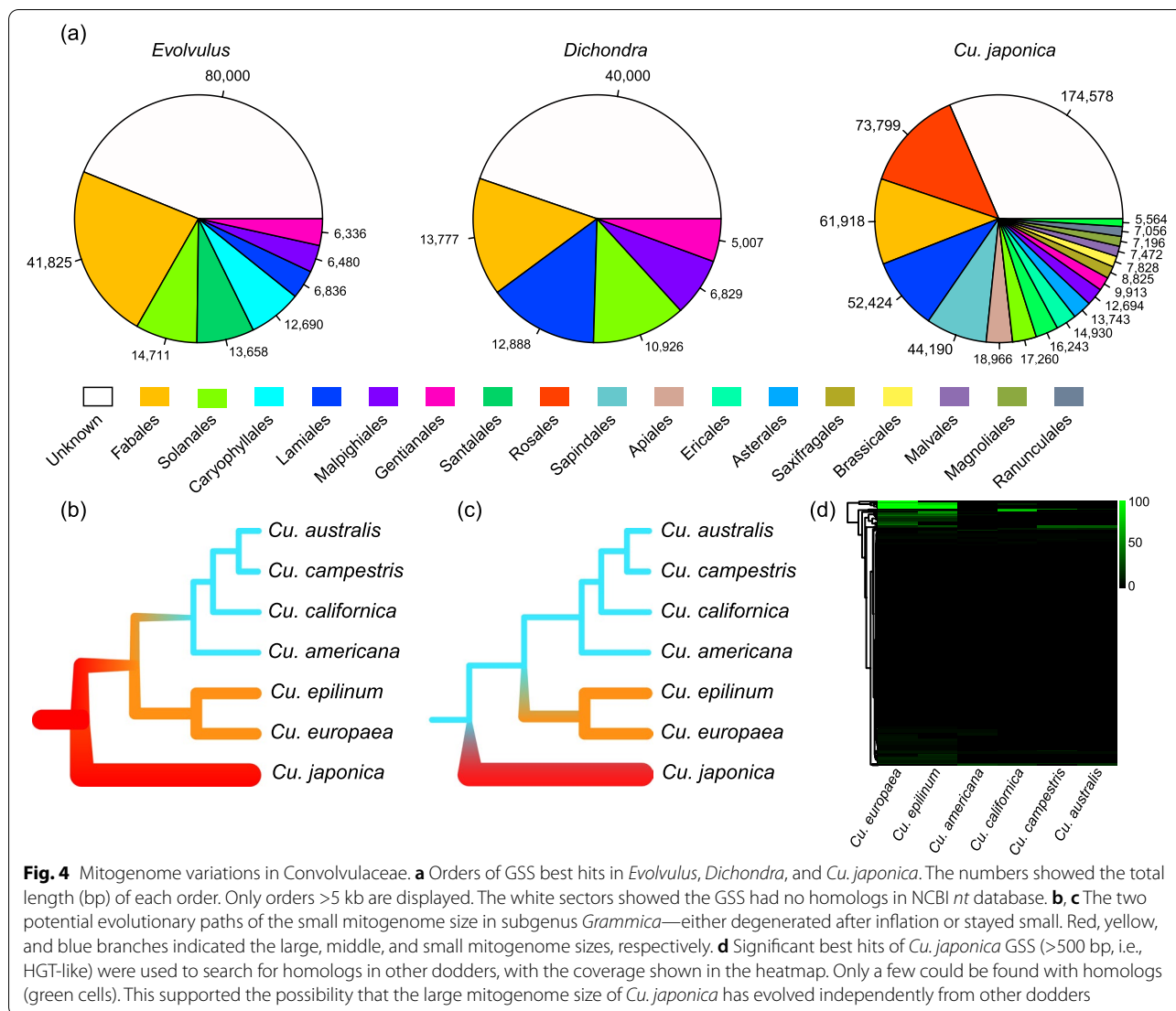


Fig. 4 Mitogenome variations in Convolvulaceae. **a** Orders of GSS best hits in *Evolvulus*, *Dichondra*, and *Cu. japonica*. The numbers showed the total length (bp) of each order. Only orders >5 kb are displayed. The white sectors showed the GSS had no homologs in NCBI *nt* database. **b, c** The two potential evolutionary paths of the small mitogenome size in subgenus *Grammica*—either degenerated after inflation or stayed small. Red, yellow, and blue branches indicated the large, middle, and small mitogenome sizes, respectively. **d** Significant best hits of *Cu. japonica* GSS (>500 bp, i.e., HGT-like) were used to search for homologs in other dodders, with the coverage shown in the heatmap. Only a few could be found with homologs (green cells). This supported the possibility that the large mitogenome size of *Cu. japonica* has evolved independently from other dodders

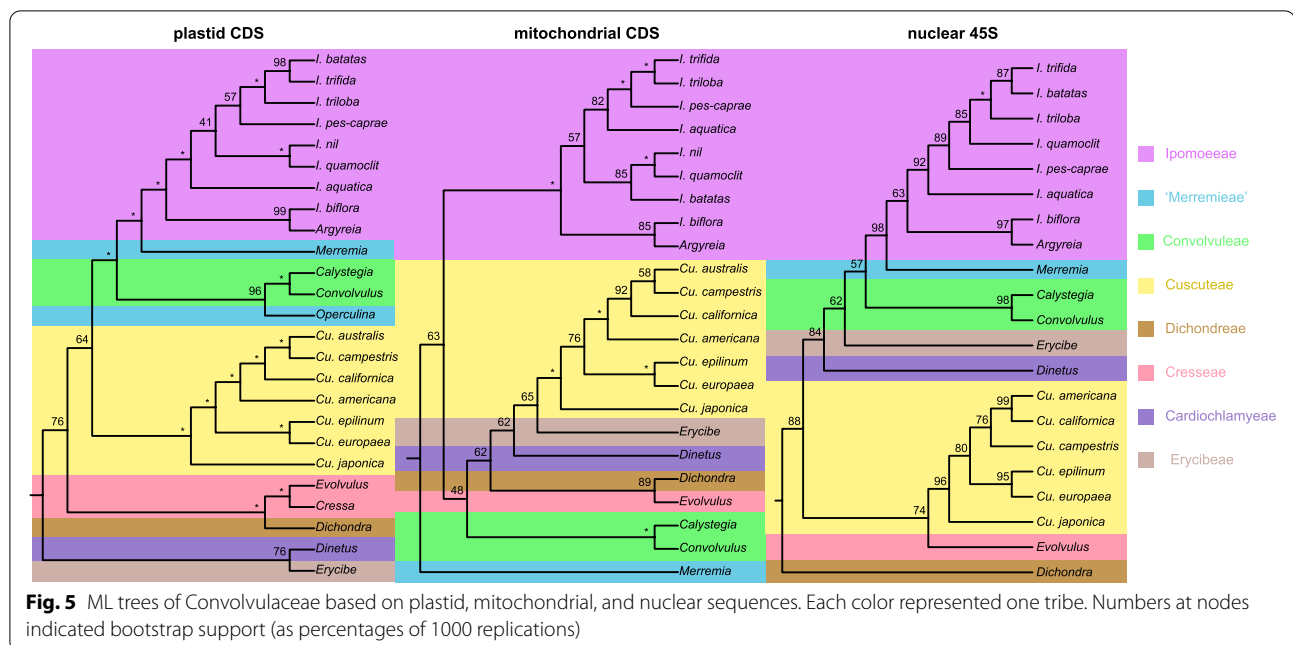
the coverage rate of each GSS. Only a few GSS yielded hits from the other dodders (Fig. 4d). It suggested that though dodders had some HGT-like sequences in common, most expansions in *Cu. japonica* occurred independently after the speciation. The mitogenome size evolved in a divergent manner in dodders. A similar situation was also observed for *Viscum scurruloideum*, which possesses the smallest mitogenome in angiosperms (only 66 kb) while its relative *V. album* outnumbers it 8.6-fold in size [45, 46]. Why some dodders and *V. scurruloideum* evolved in an opposite direction, and how they maintained a small mitogenome are intriguing questions to be answered in the future.

Phylogenetic relationships

Sequences of plastid protein coding sequences (CDS), mitochondrial CDS, and nuclear 45S (18S, 5.8S, and

25S rRNAs and the spacer regions) were employed to build the maximum-likelihood (ML) trees of Convolvulaceae (Fig. 5; Additional file 5). Plastomes of *Operculina macrocarpa* (GenBank: KF242502) and *Cressa cretica* (NC_035516) were included. The plastid matrix contained 77,711 columns with 16,689 parsimony informative sites (PIS). The plastid tree has a similar topology to previous studies [3–5]. However, the two species from “Merremieae,” *Merremia* and *Operculina*, were nested with Convolvuleae. Merremieae was suggested as polyphyletic [5, 13] and then re-classified based on several DNA markers and morphological data [1]. Our result revealed that the monophyly of Merremieae was likely still uncertain. The support of the Cuscutae clade was low, the sister group of dodders remained unresolved.

The mitochondrial matrix had 30,753 columns but only with 1033 PIS. The mitochondrial tree got a very different



topology and was poorly supported. The ability of mitochondrial sequences in phylogeny in family or order level was challenged [47]. The nuclear tree suggested a closer relationship between dodders and Erycibeae. However, the support of the nuclear tree was also weak, which may also be caused by a lack of PIS (5906 columns with 548 PIS).

All family members harbor the broken *ccmFc*

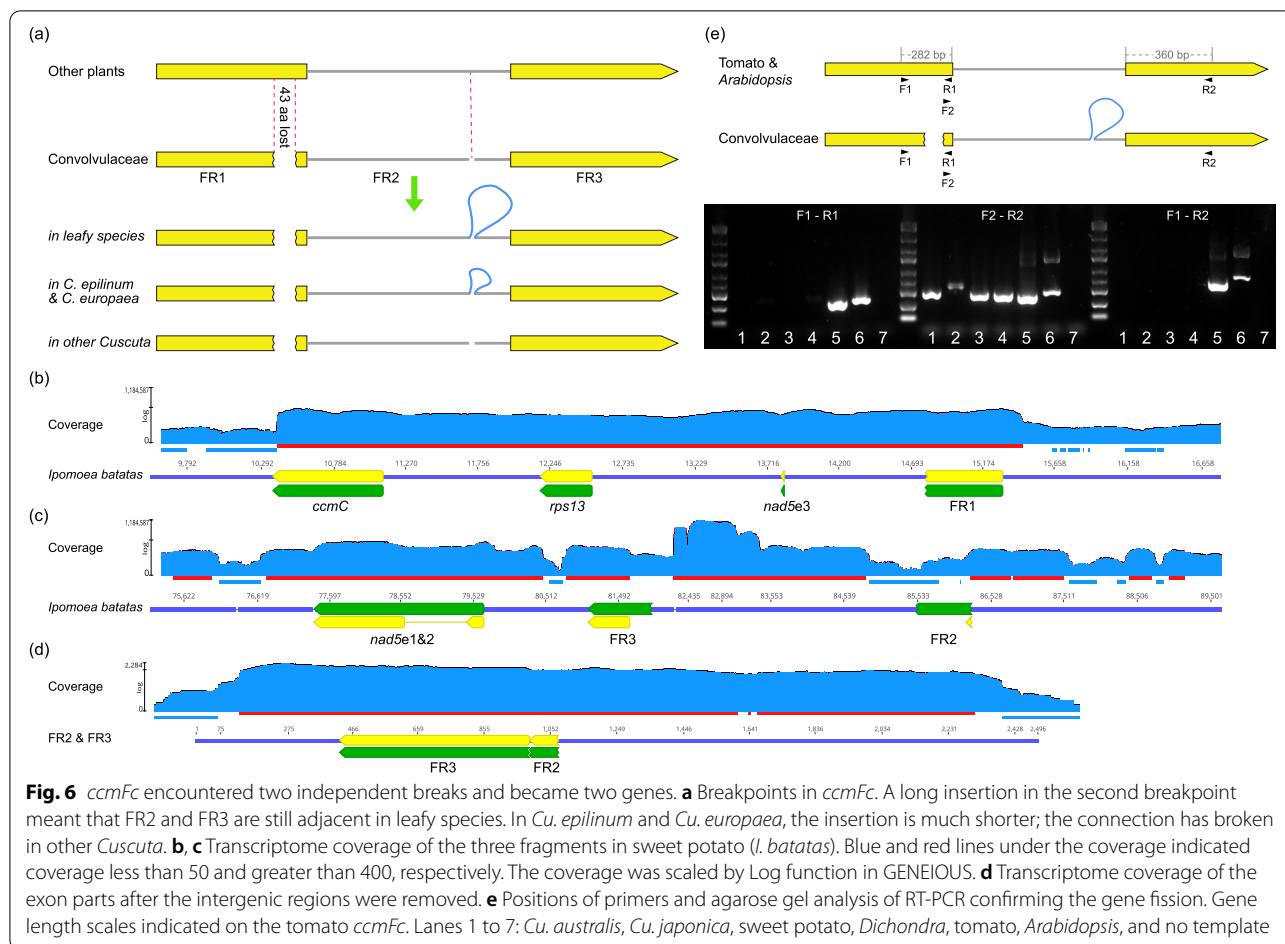
A very unusual phenomenon was observed in the *ccmFc* gene. *ccmFc* is one subunit of the cytochrome *c* maturation (CCM) system and is involved in the final stage of the maturation process [48, 49]. The *c*-type cytochrome is an essential component of the mitochondrial electron transport chain and delivers electrons between complexes III and IV. Typically, the gene contains two exons and a group II *cis*-splicing intron, and the structure and sequence of the gene are remarkably conserved in almost all angiosperms. Differently, in all members of Convolvulaceae, it has been divided into three fragments. The first break occurred close to the 3'-end of the first exon, causing a *ca.* 43 aa loss (between positions 164 to 206 in tomato); the second break occurred within the intron without base losses (Fig. 6a). Our data included multiple long sequencing reads, and we detected no large repeats around *ccmFc* fragments, arguing against incorrect assembly. Our rich sampling suggests the breaks most likely occurred to the common ancestor of this family.

The genomic distribution of the three *ccmFc* fragments varied among species (Fig. 6a). In leafy members, the first fragment (FR1) formed a new gene block together with

nad5 exon 3, *rps13*, and *ccmC* (this block had been rearranged in dodders). The second (FR2) and third (FR3) fragments were still connected in leafy species, with a *ca.* 3.5 kb insertion (2.8 kb of plastid origin) in between; in *Cu. epilinum* and *Cu. europaea*, the insertion was reduced to *ca.* 1.3 kb; while in other dodders, the connection was broken, and FR2 and FR3 had become separated.

CcmFc encountered fission and a splicing shift

Considering the fast evolution of mitochondrial non-coding regions (Fig. 3), the presence of all three *ccmFc* fragments in Convolvulaceae probably implies they are still functional. We searched for DNA and amino acid sequences of *ccmFc* in known Convolvulaceae genomes to detect potential nuclear transfers, but no high-confidence matches were obtained. We then mapped the rRNA depletion transcriptome of sweet potato to the mitogenome to check for evidence of active expression. All three fragments yielded high expression levels (Fig. 6b, c). FR1 formed a new transcription unit with the downstream *nad5* exon 3, *rps13*, and *ccmC* (Fig. 6b). No splicing sites were found around the 3'-end of FR1, while FR2 and FR3 had reads mapped to both ends (Fig. 6c, d). These results support FR1 becoming an independent gene (hereafter, *ccmFc1*), whereas FR2 and FR3 are two exons of another gene (hereafter, *ccmFc2*). To validate this, we synthesized cDNA of sweet potato, *Dichondra*, *Cu. australis*, *Cu. japonica*, tomato, and *Arabidopsis* from mRNA using reverse transcriptase (RT) and performed PCR analyses. Primer pair "F1 + R1" yielded a band in tomato and *Arabidopsis* but not Convolvulaceae species



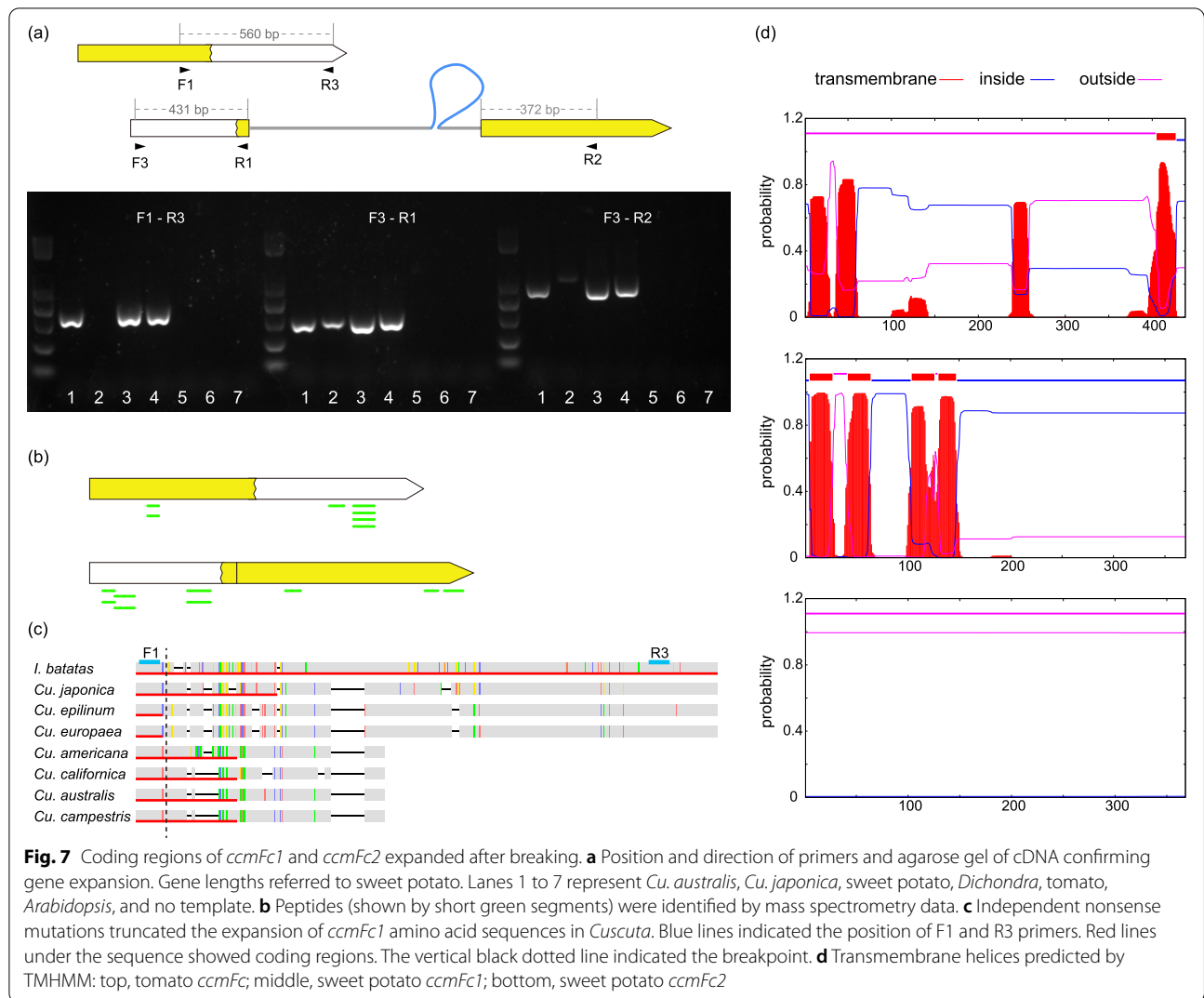
(Fig. 6e). In contrast, primer pair “F2 + R2” that should amplify a part within the region containing FR2 and FR3 yielded a band for all cDNA templates, serving as a positive control for the quality of the cDNA templates. These results support the occurrence of the fission event deduced from the transcriptome data.

The two exons of *ccmFc2* are still concatenated in most family members but are separated in some dodders (Fig. 6b). The existence of both *cis*- and *trans*-splicing in angiosperm mitogenomes raises the obvious question of how *ccmFc2* mRNA is spliced. The following evidence indicated splicing could be *trans* in all species. On the one hand, the unspliced intron-containing pre-mRNA could be cloned since random primers were used for reverse transcription. Another band fainter in brightness was observed in the gel for tomato and *Arabidopsis*, but not for Convolvulaceae species (Fig. 6e). On the other hand, a *cis*-splicing intron might also be transcribed at a high level (e.g., *cis*-splicing intron of *nad5* in Fig. 6c). However, the transcriptomic mapping of sweet potato displayed a relatively low coverage close to FR2 (Fig. 6c, blue line). Similar results were also obtained from the

very recent work of the two dodders *Cu. australis* and *Cu. campestris* [50].

Coding sequences expanded after fission

Along with the fission, there was an increase in indels and divergent bases in the coding regions of *ccmFc*, which accelerated the substitution rates (see Additional file 3, *ccmFc1* and *ccmFc2*). Additionally, after the break, FR1 was separated from the stop codon. Thus, it required a new stop codon downstream; similarly, FR2 required a new initiation codon upstream. However, we determined that the closest stop and ATG initiation codons were far away, resulting in a *ca.* 185 and 118 aa expansion of *ccmFc1* and *ccmFc2*, respectively. New RNA editing sites were absent in the expanded genic regions. We cloned cDNA to confirm the gene expansion (Fig. 7a). The expansions were further confirmed by the sweet potato mass spectrometry data, with multiple hits detected during peptide identification (Fig. 7b). In dodders, *ccmFc1* gained independent nonsense mutations shortening its length back to the breakpoint (Fig. 7c).



CcmFc is localized to the mitochondrial inner membrane and contains multiple predicted transmembrane helices in both exons (Fig. 7d [48]). However, in sweet potato, our predictions for *ccmFc1* and *ccmFc2* (performed by TMHMM Server 2.0 [51]) revealed four transmembrane helices in *ccmFc1* but none in *ccmFc2* (Fig. 7d). The breaks and expansions might have altered the protein structure.

Cis- and trans-splicing may not differ much

The mechanism of the shift from *cis*- to *trans*-splicing remains unclear. What's This Factor 9 (WTF9), a nuclear-encoded plant organelle RNA recognition protein, can bind directly to the 48 nucleotides of the *ccmFc* intron to regulate *cis*-splicing [49, 52]. The molecular chaperone heat shock protein 60 (HSP60)

interacts with WTF9 during this process [52]. To detect potential changes caused by the *trans*-splicing shift, we used *Arabidopsis* WTF9 and HSP60 as references to search for homologs in genomes of Convolvulaceae and other angiosperms. WTF9 is retained in all species of Convolvulaceae and even duplicated in *Ipomoea* (Fig. 8a). We checked the expression level of WTF9 in *I. trifida*, *I. triloba*, and two species of Solanaceae, potato and *S. pimpinellifolium*. The expression pattern of *Ipomoea* WTF9-1 resembled that of *Solanum* (Fig. 8b). A similar situation was observed for HSP60 (Additional file 6). After the splicing shift, the two intronic fragments were almost entirely preserved in all species, including the 48-nucleotide binding region (except *Cu. americana*, which has a 28 bp insertion in the binding region). Based on these,

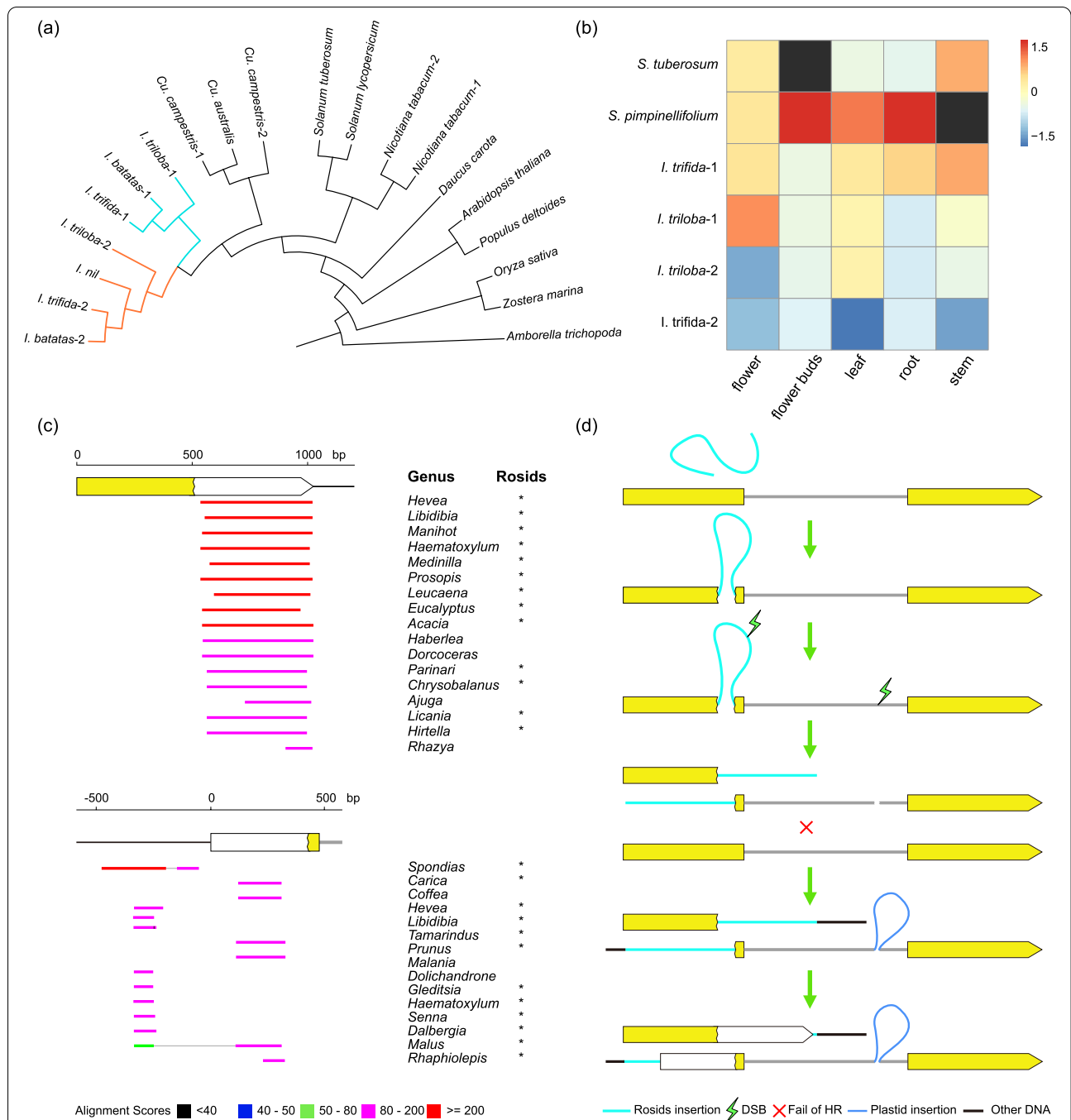


Fig. 8 Phylogeny and expression of *WTF9* and a potential evolutionary history of *ccmFc* in Convolvulaceae. **a** Phylogeny of *WTF9* in angiosperms. Colored branches in *Ipomoea* showed duplication. **b** *WTF9* expression level in *Solanum* and *Ipomoea*. Black boxes indicated missing data. **c** BLASTN searches of the downstream of FR1 and upstream of FR2 against NCBI nt database. Top hits were mainly from Rosids (*) after masked hits from Convolvulaceae and other parasitic plants. **d** A potential evolutionary trajectory of *ccmFc* in Convolvulaceae

we speculated that the mechanism between *cis*- and *trans*-splicing might be similar, but further research is needed to test this hypothesis.

Variations in Convolvulaceae *CcmFc* is unprecedented among angiosperms

Under selection pressure and homologous recombination (HR) repair, plant mitochondrial genes evolve more slowly than other genome compartments [53, 54]. The mitogenomic structure in angiosperms is exceptionally dynamic, while gene sequences and structure, as well as methods of splicing, are well conserved [55, 56]. *Trans*-splicing has evolved independently several times in land plants, such as in lycophytes and gymnosperms [57–60]. In angiosperms, the emergence of *trans*-splicing was less frequent. Aside from *nad1*, *nad2*, and *nad5*, only *cox2* in *Allium* was reported to undergo *trans*-splicing [61–64]. Our discovery of *ccmFc* in Convolvulaceae represents a novel type of *trans*-splicing in angiosperms. Gene fission has also occurred several times in plant mitogenomes, such as *ccmFc* in *Marchantia* and *ccmFn* in Brassicaceae, *Allium*, and *Trifolium*, caused either by deletion-based frameshift mutation or rearrangements [65–68]. As far as we know, fission followed by expansions found in this study is the first time reported in mitogenomes. Among them, the *ccmFc* in Convolvulaceae might be the most complicated, integrating gene fission, splicing shift, and expansions. We also searched the downstream of FR1 and upstream of FR2 against the NCBI *nt* database, finding that both were most similar to the mitochondrial sequences of rosids after masked hits of Convolvulaceae (Fig. 8c; Additional file 7). We hypothesized that a long mitochondrial sequence from rosids inserted into the *ccmFc* gene of Convolvulaceae common ancestor occasionally (Fig. 8d). Then double-strand breaks (DSBs [69]); occurred inside the insertion and the intron. The long insertion could make it difficult to repair the mutation through HR. Imprecise non-homologous repair finally created the gene fission and brought the plastid insertion inside the intron. The latter may have interrupted the standard secondary structure of the group II intron, resulting in a shift from *cis*-splicing to *trans*. HGTs are widespread among mitogenomes but mostly non-functional [70, 71]. In some cases, HGT can replace native genes completely or partially [41, 72]. The chimeric *ccmFc1* and *ccmFc2* in Convolvulaceae fuse novel HGT and native coding sequences, representing a novel manner of how HGT functionally impacts the organism.

An immediate question is how these changes would modify the structure of the CCM complex or even the efficiency of oxidative phosphorylation. The elevated

substitution rates and potential changes in transmembrane helices may have affected the protein structure. Recent studies have shown that modifying cytochrome *c* maturation can optimize extracellular electron transfer for higher electron flux in engineered *Escherichia coli* [73]. Our results provide an example of plant respiratory diversity, which may have implications for the engineering of plant respiration in the future.

Conclusions

The family Convolvulaceae is famous for its extensive uses, diverse morphologies, and different trophic modes. In this study, our family-scale analyses revealed that the same amazement also exists in their genetic materials. Plastids (including chloroplasts) and mitochondria play crucial roles in cellular energy supply and metabolism. Normally the structure and genes of plastome and coding sequences of mitogenome are well-conserved. However, it has some changes in Convolvulaceae. Even in green leafy members, the IR boundaries are all atypical, and uncommon features such as intron losses, nuclear transfer (*infA*), duplications, and foreign insertions also exist. The mitogenome of *Cu. japonica* was significantly inflated by HGT-like sequences, whereas other dodders preserved in a small size like their leafy relatives—the mitogenome evolves divergently in dodders. Besides, dodder mitochondrial genes have more losses and higher substitution rates, but no HGT events. The most notable variation is the mitochondrial *ccmFc*. It is highly conserved in other plants but likely encountered gene fission, splicing shift, and coding expansion in the common ancestor of Convolvulaceae, which is unique in angiosperms. Nuclear genes involved in the splicing process seems still functional, raising the question of what differs between *cis*- and *trans*-splicing on mechanism. These unusual changes of *ccmFc* were likely caused by HGT insertion followed by double-strand breaks and imprecise repairs. Our data provide valuable genetic resources for studying the evolution and phylogeny of Convolvulaceae and plant parasitism. The results presented in this study are also helpful to understand the diversification of mitochondrial complexes and gene innovation.

Methods

Sampling and sequencing

Accessions of 12 Convolvulaceae species were collected from different places in China (Additional file 8). Sequencing libraries were prepared using a NEB Next® Ultra DNA Library Prep Kit, then sequenced on the Illumina HiSeq 4000 platform to generate PE150 reads. *Dichondra micrantha* was also sequenced using the Oxford Nanopore promethION platform. Raw sequencing data for the other 10 species were obtained from

NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>). Details of the reads are provided in Additional file 9.

Genome assembly and annotation

Short reads were filtered using TRIMMOMATIC v0.36 [74]. Plastomes were assembled using GetOrganelle [75] and annotated using *I. nil* (GenBank: NC_031159) as the reference. The assembly of mitogenomes was as described in our previous work [47]. Briefly, de novo assemblies of short and long reads were performed using SPAdes v3.13.1 [76] and FLYE v2.8.3 [77], respectively. Then mitochondrial contigs were identified from total contigs by BLASTN against the *I. nil* mitogenome. These contigs were connected and/or extended manually in GENEIOUS R10 (Biomatters, Inc.) by mapping reads back and checking both ends. Long repeats and plastid insertions were the two major challenges of the assembly. Repeat regions were resolved using short-read sequencing coverage; plastid insertions were resolved based on the locations of their plastid counterparts. The mitogenome could get one or several circles after the repeats and plastid insertions were properly determined. *Cu. japonica* yielded one linear chromosome since repeats at both ends could not be connected. The four dodders from subgenus *Grammica* (*Cu. americana*, *Cu. australis*, *Cu. californica*, and *Cu. campestris*) only obtained draft mitogenomes because of the large amounts of repeats. Putative mitochondrial protein-coding and rRNA genes were annotated based on similarity to known mitochondrial genes. tRNAs were predicted using tRNAscan-SE v2.0 [78]. Plastid insertion and repeat length were determined by BLASTN v2.10.0+ [79] against Convolvulaceae plastomes (identity > 90%) and itself (identity > 95%), respectively. Common sequences were determined initially by BLASTN (word size 16, *e*-value 1e-5), and then counted using custom scripts.

Analyses of genome synteny and Genus-specific sequences

Plastome (IR removed) and mitogenome syntenies were plotted using Python version MCscan of JCVI 192 utility libraries v1.1.17 [80]. The DNA sequences were first divided into 50 bp fragments, and these fragments were then forced to be used as “genes” to search for homologous regions. Genus-specific sequences (GSS) meant sequences that shared no homology with other genera. For species in *Ipomoea* and *Cuscuta*, GSS also included the species-specific sequences. GSS were identified by searching each mitogenome against other Convolvulaceae mitogenomes, with *e*-value of 1e-5 and word size of 16. The GSS of *Cu. japonica*, *Dichondra*, and *Evolvulus* were searched against NCBI *nt* database to find the most potential donors, with only hits longer than 100

bp considered. Subsequently, the best matches were 184 grouped into orders (Additional file 3). Orders with total length >5 kb were plotted as pie charts in R (<https://www.r-project.org/>). To test whether other dodders have experienced expansion, high-confidence HGT-like sequences of *Cu. japonica* (best hits >500 bp) were searched against other dodders, with the percentage of the coverage calculated. Finally, the heatmaps were generated using PHEATMAP (<https://github.com/raivo kolde/pheatmap>) in R.

Phylogenetic analyses

Potential HGT events of the mitochondrial genes were detected. CDS of Convolvulaceae and another 42 seed plants were aligned using MAFFT v7 [81] with the “auto” mode. Then ML trees were built using IQTREE v1.6.12 [82] with parameter *-bb 1000 -m GTR+G4+F-me 0.0001 runs 5*. The trees were plotted using FIGTREE v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). GenBank accessions were listed behind the species on the trees (Additional file 3).

The plastid CDS were obtained from our assemblies and *Ipomoea batatas* (GenBank: NC_026703), *I. nil* (NC_031159), *I. trifida* (NC_034670), *I. triloba* (NC_037913), *Cressa cretica* (NC_035516), *Cuscuta australis* (NC_045885), *Cu. campestris* (NC_052920), *Convolvulus arvensis* (MW054627), *Operculina macropcarpa* (KF242502), *Nicotiana tabacum* (Z00044), *Solanum melongena* (MN218080), and *Solanum lycopersicum* (NC_007898). Mitochondrial CDS were obtained from our assemblies and *I. nil* (AP017303), *N. tabacum* (NC_006581), *S. lycopersicum* (NC_035963), and *S. melongena* (NC_050334). RNA editing sites were predicted using the PREP website (Mower, 2009) and removed manually. The nuclear 45S were obtained from *de novo* contigs and *Nicotiana benthamiana* (KP824745). Aligned plastid and mitochondrial (*rps7* was excluded from an HGT) CDS were concatenated into one matrix, respectively. Then the three datasets (plastid CDS, mitochondrial CDS, and nuclear 45S) were built the ML trees as mentioned above. Solanaceae were always used as outgroups.

Peptide identification from published proteomics data

Protein mass spectrometry data for sweet potato (*I. batatas*) were downloaded from ProteomeXchange PXD012999 [83]. A Trans-Proteomic Pipeline (TPP) v5.2.0 [84] comet search was used for peptide identification with default parameters.

Gene family analyses

Amino acid sequences of *Arabidopsis thaliana* WTF9 (AT3G24320), HSP60-2 (AT2G33210), HSP60-3A

(AT3G13860), and *HSP60-3B* (AT3G23990) were used as references to search for homologs in Convolvulaceae, including sweet potato [85], *I. trifida* and *I. triloba* [86], *I. nil* [15], *Cuscuta australis* [19], and *Cu. campestris* [20], and other angiosperms, including *Nicotiana tabacum* [87], *Amborella trichopoda* v1.0 [88], *Daucus carota* v2.0 [89], *Populus trichocarpa* v3.0 [90], *Oryza sativa* v7 [91], tomato *Solanum lycopersicum* iTAG2.4 [92], potato *Solanum tuberosum* v4.03 [93], and *Zostera marina* v2.2 [94] (the last seven were downloaded from Phytozome v13: <https://phytozome.jgi.doe.gov/>). The amino acid sequences were then aligned through MAFFT, and trees were then reconstructed using IQTREE (parameter: *-bb 1000 -m LG+F+G4 runs 10 -me 0.0001*).

The expression level data of *Ipomoea trifida* and *I. triloba* were obtained from Sweetpotato Genomics Resource [86]; those of *Solanum pimpinellifolium* were from the Tomato Functional Genomics Database [95], and those of potato were from SpudDB [96]. Heatmaps were plotted using PHEATMAP in R (<https://r-charts.com/correlation/pheatmap/>) by normalizing each gene with Z-score.

Expression examination

Transcriptomes enriched with oligo-dT would introduce bias since organelle transcripts do not generally contain a polyA tail [97]. Based on that, sweet potato rRNA-depletion transcriptome data (SRA: SRR10538086 [98]) were employed and were mapped to the mitogenome to check the expression.

Fresh leaves of sweet potato, *Dichondra*, tomato, *Arabidopsis*, and fresh vines of *Cuscuta australis* and *Cu. japonica* were used for RNA extraction using an Omega RNA isolation kit and reverse-transcribed into cDNA using random primers. Primers used for RT-PCR were designed using GENEIOUS (Additional file 10).

Abbreviations

IR: Inverted repeat; LSC: Long single-copy; SSC: Short single-copy; HGT: Horizontal gene transfer; GSS: Genus- or/and species-specific sequences; CDS: Coding sequences; ML: Maximum-likelihood; PIS: Parsimony informative sites; CCM: Cytochrome c maturation; FR1: The first fragment; FR2: The second fragment; FR3: The third fragment; RT: Reverse transcriptase; WTF9: What's This Factor 9; HSP60: Heat shock protein 60; HR: Homologous recombination; DSBs: Double-strand breaks.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01250-1>.

Additional file 1. Plastome content. The black, grey and blank cells represent intact, pseudo and missing, respectively. Coding genes with disrupted reading frames, premature stop codons, or non-triplet frameshifts were annotated as pseudogenes.

Additional file 2. Mitogenome content. The black, grey and blank cells represent intact, pseudo and missing, respectively. Coding genes with disrupted reading frames, premature stop codons, or non-triplet frameshifts were annotated as pseudogenes.

Additional file 3. Most mitochondrial genes in dodders.

Additional file 4. BLASTN results of *Cu. japonica* genus- and species-specific sequences.

Additional file 5. Plastid CDS, Mitochondrial CDS, and Nuclear 45S.

Additional file 6. Similar situation observed for *HSP60*.

Additional file 7. BLASTN searches of the downstream of FR1 and upstream of FR2. *Erycibe* was used as the query. Parameter: word_size 16, e-value 1e-5.

Additional file 8. Collections of materials.

Additional file 9. Data used for organelle genome assemblies.

Additional file 10. Primers used to confirm gene fission and expansion.

Acknowledgements

We gratefully acknowledge Liansheng Xu and Pengwei Li (South China Botanical Garden, The Chinese Academy of Sciences) and Chang An (Guangxi University) for helping collect materials and providing photographs.

Authors' contributions

Y.F. designed the project. Y.L. and P.L. collected the materials and did the sequencing. Y.L., P.L., Y.Z., D.A., R.P., Z.F., M.H., X.L., and Y.F. performed the research and wrote the manuscript. The authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (32100168), the Chinese State Key Laboratory of Subtropical Silviculture (KF202001), the National Natural Science Foundation of China (31970225), the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2019R01002), the Natural Science Foundation of Fujian Province of China (2021J05061), and the Scientific Research Foundation for the High-level Talents, Fujian University of Traditional Chinese Medicine (X2020012-talents).

Availability of data and materials

Raw reads have been deposited on NCBI under BioProject PRJNA737311 and accession SRR14812959 - SRR14812971 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA737311>) [99]. The assembled mitogenomes and plastomes have been deposited to NCBI under accessions MZ240723 - MZ240750 and BK059236 - BK059244 (TPA) and CNGBdb under project CNP0001927 and accessions N_000011218 - N_000011230 and N_000011273 - N_000011309 (<https://db.cngb.org/search/project/CNP0001927>). The used scripts can be found in Github (<https://github.com/fengyanlei33/Convolvulaceae-mitogenome-project>) [100].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Pharmacy, Fujian University of Traditional Chinese Medicine, Fuzhou 350122, Fujian, China. ²MOE Laboratory of Biosystem Homeostasis and Protection, College of Life Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China. ³Institute of Crop Science, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, Zhejiang,

China. ⁴Department of Genetics and Plant Breeding, Sylhet Agricultural University, Sylhet Division 3100, Sylhet, Bangladesh. ⁵ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 310027, China. ⁶College of Life Science, Sichuan Normal University, Chengdu 610101, Sichuan, China. ⁷Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310024, Zhejiang, China. ⁸Institute of Biology, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang, China.

Received: 10 October 2021 Accepted: 8 February 2022

Published online: 16 February 2022

References

1. Simoes AR, Staples G. Dissolution of Convolvulaceae tribe Merremieae and a new classification of the constituent genera. *Bot J Linn Soc.* 2017;183(4):561–86. <https://doi.org/10.1093/botlinnean/box007>.
2. Stevens PF. Angiosperm Phylogeny Website. Version 14, 2017. 2001 onwards. <http://www.mobot.org/MOBOT/research/APweb/>.
3. Stefanovic S, Krueger L, Olmstead RG. Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *Am J Bot.* 2002;89(9):1510–22. <https://doi.org/10.3732/ajb.89.9.1510>.
4. Stefanovic S, Olmstead RG. Testing the phylogenetic position of a parasitic plant (*Cuscuta*, Convolvulaceae, asteridae): Bayesian inference and the parametric bootstrap on data drawn from three genomes. *Syst Biol.* 2004;53(3):384–99. <https://doi.org/10.1080/10635150490445896>.
5. Stefanovic S, Austin D, Olmstead R. Classification of Convolvulaceae: a phylogenetic approach. *Syst Bot.* 2009;28:791–806. <https://doi.org/10.1043/02-45.1>.
6. Eserman LA, Tiley GP, Jarret RL, Leebens-Mack JH, Miller RE. Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convolvulaceae) based on whole plastome sequences. *Am J Bot.* 2014;101(1):92–103. <https://doi.org/10.3732/ajb.1300207>.
7. Park I, Yang S, Kim WJ, Noh P, Lee HO, Moon BC. The complete chloroplast genomes of six ipomoea species and indel marker development for the discrimination of authentic pharbitidis semen (seeds of *I. nil* or *I. purpurea*). *Front. Plant Sci.* 2018;9:965. <https://doi.org/10.3389/fpls.2018.00965>.
8. Sun J, Dong X, Cao Q, Xu T, Zhu M, Sun J, et al. A systematic comparison of eight new plastome sequences from *Ipomoea* L. *PeerJ.* 2019;7:e6563. <https://doi.org/10.7717/peerj.6563>.
9. Funk HT, Berg S, Krupinska K, Maier UG, Krause K. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 2007;7:45. <https://doi.org/10.1186/1471-2229-7-45>.
10. McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 2007;7. <https://doi.org/10.1186/1471-2229-7-57>.
11. Braukmann T, Kuzmina M, Stefanovic S. Plastid genome evolution across the genus *Cuscuta* (Convolvulaceae): two clades within subgenus *Grammica* exhibit extensive gene loss. *J Exp Bot.* 2013;64(4):977–89. <https://doi.org/10.1093/jxb/ers391>.
12. Banerjee A, Stefanović S. Reconstructing plastome evolution across the phylogenetic backbone of the parasitic plant genus *Cuscuta* (Convolvulaceae). *Bot J Linn Soc.* 2020;194(4):423–38. <https://doi.org/10.1093/botlinnean/boaa056>.
13. Simoes AR, Culham A, Carine M. Resolving the unresolved tribe: a molecular phylogenetic framework for the Merremieae (Convolvulaceae). *Bot J Linn Soc.* 2015;179(3):374–87. <https://doi.org/10.1111/boj.12339>.
14. Shidhi PR, Biju VC, Anu S, Vipin CL, Deelip KR, Achuthsankar SN. Genome characterization, comparison and phylogenetic analysis of complete mitochondrial genome of *evolvulus alsinoides* reveals highly rearranged gene order in solanales. *Life (Basel).* 2021;11(8). <https://doi.org/10.3390/life11080769>.
15. Hoshino A, Jayakumar V, Nitasaka E, Toyoda A, Noguchi H, Itoh T, et al. Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. *Nat Commun.* 2016;7:13295. <https://doi.org/10.1038/ncomms13295>.
16. Wicke S, Muller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, et al. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell.* 2013;25(10):3711–25. <https://doi.org/10.1105/tpc.113.113373>.
17. Feng YL, Wicke S, Li JW, Han Y, Lin CS, Li DZ, et al. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol Evol.* 2016;8(7):2164–75. <https://doi.org/10.1093/gbe/ewv144>.
18. Revill MJ, Stanley S, Hibberd JM. Plastid genome structure and loss of photosynthetic ability in the parasitic genus *Cuscuta*. *J Exp Bot.* 2005;56(419):2477–86. <https://doi.org/10.1093/jxb/eri240>.
19. Sun G, Xu Y, Liu H, Sun T, Zhang J, Hettenhausen C, et al. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat Commun.* 2018;9(1):2683. <https://doi.org/10.1038/s41467-018-04721-8>.
20. Vogel A, Schwacke R, Denton AK, Usadel B, Hollmann J, Fischer K, et al. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat Commun.* 2018;9(1):2515. <https://doi.org/10.1038/s41467-018-04344-z>.
21. McNeal JR, Kuehl JV, Boore JL, Leebens-Mack J, dePamphilis CW. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *Plos One.* 2009;4(6):e5982. <https://doi.org/10.1371/journal.pone.0005982>.
22. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell.* 2001;13(3):645–58. <https://doi.org/10.1105/tpc.13.3.645>.
23. Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biol.* 2012;12(1):61. <https://doi.org/10.1186/1471-2229-12-61>.
24. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst Bot.* 2015;40(1):336–51. <https://doi.org/10.1600/036364415x686620>.
25. Spooner DM, Ruess H, Iorizzo M, Senalik D, Simon P. Entire plastid phylogeny of the carrot genus (*Daucus*, Apiaceae): concordance with nuclear data and mitochondrial and nuclear DNA insertions to the plastid. *Am J Bot.* 2017;104(2):296–312. <https://doi.org/10.3732/ajb.1600415>.
26. Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biol Evol.* 2013;5(10):1872–85. <https://doi.org/10.1093/gbe/evt140>.
27. Rabah SO, Lee C, Hajrah NH, Makki RM, Alharby HF, Alhebshi AM, et al. Plastome sequencing of ten nonmodel crop species uncovers a large insertion of mitochondrial DNA in cashew. *Plant. Genome.* 2017;10(3). <https://doi.org/10.3835/plantgenome2017.03.0020>.
28. Gandini CL, Sanchez-Puerta MV. Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. *Sci Rep.* 2017;7:43402. <https://doi.org/10.1038/srep43402>.
29. Ma PF, Zhang YX, Guo ZH, Li DZ. Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. *Sci Rep.* 2015;5:11608. <https://doi.org/10.1038/srep11608>.
30. Saarela JM, Wysocki WP, Barrett CF, Soreng RJ, Davis JI, Clark LG, et al. Plastid phylogenomics of the cool-season grass subfamily: clarification of relationships among early-diverging tribes. *AoB Plants.* 2015;7. <https://doi.org/10.1093/aobpla/plv046>.
31. Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC Evol Biol.* 2015;15:50. <https://doi.org/10.1186/s12862-015-0321-5>.
32. Burke SV, Ungerer MC, Duvall MR. Investigation of mitochondrial-derived plastome sequences in the *Paspalum* lineage (Panicoideae; Poaceae). *BMC Plant Biol.* 2018;18(1):152. <https://doi.org/10.1186/s12870-018-1379-1>.

33. Westwood JH, Yoder JI, Timko MP, dePamphilis CW. The evolution of parasitism in plants. *Trends Plant Sci.* 2010;15(4):227–35. <https://doi.org/10.1016/j.tplants.2010.01.004>.
34. Nickrent DL. Parasitic angiosperms: how often and how many? *Taxon.* 2020;69(1):5–27. <https://doi.org/10.1002/tax.12195>.
35. Yoshida S, Kim S, Wafula EK, Tanskanen J, Kim YM, Honaas L, et al. Genome sequence of *striga asiatica* provides insight into the evolution of plant parasitism. *Curr Biol.* 2019;29(18):3041–3052 e3044. <https://doi.org/10.1016/j.cub.2019.07.086>.
36. Cai L, Arnold BJ, Xi Z, Khost DE, Patel N, Hartmann CB, et al. Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). *Curr Biol.* 2021;31(5):1002–1011 e1009. <https://doi.org/10.1016/j.cub.2020.12.045>.
37. Lyko P, Wicke S. Genomic reconfiguration in parasitic plants involves considerable gene losses alongside global genome size inflation and gene births. *Plant Physiol.* 2021. <https://doi.org/10.1093/plphys/kiab192>.
38. Wicke S, Naumann J. Molecular evolution of plastid genomes in parasitic flowering plants. In: Chaw S-M, Jansen RK, editors. *Advances in botanical research*: Academic Press; 2018. p. 315–47. <https://doi.org/10.1016/bs.abr.2017.11.014>.
39. Bellot S, Cusimano N, Luo S, Sun G, Zarre S, Groger A, et al. Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasite family Cynomoriaceae in the Saxifragales. *Genome Biol Evol.* 2016;8(7):2214–30. <https://doi.org/10.1093/gbe/evw147>.
40. Cusimano N, Renner SS. Sequential horizontal gene transfers from different hosts in a widespread Eurasian parasitic plant, *Cynomorium cocconium*. *Am J Bot.* 2019;106(5):679–89. <https://doi.org/10.1002/ajb2.1286>.
41. Sanchez-Puerta MV, Garcia LE, Wohlfeiler J, Ceriotti LF. Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. *New Phytol.* 2017;214(1):376–87. <https://doi.org/10.1111/nph.14361>.
42. Sanchez-Puerta MV, Edera A, Gandini CL, Williams AV, Howell KA, Nevill PG, et al. Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). *Mol Phylogenet Evol.* 2018;132:243–50. <https://doi.org/10.1016/j.ympev.2018.12.006>.
43. Petersen G, Cuenca A, Moller IM, Seberg O. Massive gene loss in mistletoe (*Viscum, Viscaceae*) mitochondria. *Sci Rep.* 2015;5:17588. <https://doi.org/10.1038/srep17588>.
44. Choi K-S, Park S. Complete plastid and mitochondrial genomes of *aeginetia indica* reveal intracellular gene transfer (IGT), horizontal gene transfer (HGT), and cytoplasmic male sterility (CMS). *Int J Mol Sci.* 2021;22(11). <https://doi.org/10.3390/ijms22116143>.
45. Skippington E, Barkman TJ, Rice DW, Palmer JD. Comparative mitogenomics indicates respiratory competence in parasitic *Viscum* despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. *BMC Plant Biol.* 2017;17(1):49. <https://doi.org/10.1186/s12870-017-0992-8>.
46. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A.* 2015;112(27):E3515–24. <https://doi.org/10.1073/pnas.1504491112>.
47. Feng Y, Xiang X, Akhter D, Pan R, Fu Z, Jin X. Mitochondrial phylogenomics of fagales provides insights into plant mitogenome mosaic evolution. *Front Plant Sci.* 2021;12. <https://doi.org/10.3389/fpls.2021.762195>.
48. 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. <https://doi.org/10.1074/jbc.M802621200>.
49. des Francs-Small CC, Kroeger T, Zmudjak M, Ostersetzer-Biran O, Rahimi N, Small I, et al. A PORR domain protein required for rpl2 and ccmFC intron splicing and for the biogenesis of c-type cytochromes in *Arabidopsis* mitochondria. *Plant J.* 2012;69(6):996–1005. <https://doi.org/10.1111/j.1365-3113X.2011.04849.x>.
50. Anderson BM, Krause K, Petersen G. Mitochondrial genomes of two parasitic *Cuscuta* species lack clear evidence of horizontal gene transfer and retain unusually fragmented ccmFC genes. *BMC Genomics.* 2021;22(1):816. <https://doi.org/10.1186/s12864-021-08105-z>.
51. Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics.* 2001;17(7):646–53. <https://doi.org/10.1093/bioinformatics/17.7.646>.
52. Hsu YW, Juan CT, Wang CM, Jauh GY. Mitochondrial heat shock protein 60s interact with what's this factor 9 to regulate RNA splicing of ccmFC and rpl2. *Plant Cell Physiol.* 2019;60(1):116–25. <https://doi.org/10.1093/pcp/pcy199>.
53. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A.* 1987;84(24):9054–8. <https://doi.org/10.1073/pnas.84.24.9054>.
54. Palmer JD, Herbon LA. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol.* 1988;28(1-2):87–97. <https://doi.org/10.1007/BF02143500>.
55. Cole LW, Guo W, Mower JP, Palmer JD. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Mol Biol Evol.* 2018;35(11):2773–85. <https://doi.org/10.1093/molbev/msy176>.
56. Mower JP. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion.* 2020;53:203–13. <https://doi.org/10.1016/j.mito.2020.06.002>.
57. Grewe F, Viehoveer P, Weisshaar B, Knoop V. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 2009;37(15):5093–104. <https://doi.org/10.1093/nar/gkp532>.
58. Hecht J, Grewe F, Knoop V. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol.* 2011;3:344–58. <https://doi.org/10.1093/gbe/evr027>.
59. Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, et al. Ginkgo and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol Biol Evol.* 2016;33(6):1448–60. <https://doi.org/10.1093/molbev/msw024>.
60. Guo W, Zhu A, Fan W, Adams RP, Mower JP. Extensive shifts from cis- to trans-splicing of gymnosperm mitochondrial introns. *Mol Biol Evol.* 2020;37(6):1615–20. <https://doi.org/10.1093/molbev/msaa029>.
61. Malek O, Knoop V. Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA (New York, NY).* 1998;4(12):1599–609.
62. Qiu YL, Palmer JD. Many independent origins of trans splicing of a plant mitochondrial group II intron. *J Mol Evol.* 2004;59(1):80–9. <https://doi.org/10.1007/s00239-004-2606-y>.
63. Kim S, Yoon MK. Comparison of mitochondrial and chloroplast genome segments from three onion (*Allium cepa* L.) cytoplasm types and identification of a trans-splicing intron of *cox2*. *Curr Genet.* 2010;56(2):177–88. <https://doi.org/10.1007/s00294-010-0290-6>.
64. Kim S, Bang H, Patil BS. Origin of three characteristic onion (*Allium cepa* L.) mitochondrial genome rearrangements in *Allium* species. *Sci Hortic.* 2013;157:24–31. <https://doi.org/10.1016/j.scienta.2013.04.009>.
65. Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, et al. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol.* 1992;223(1):1–7.
66. Handa H, Bonnard G, Grienenberger JM. The rapeseed mitochondrial gene encoding a homologue of the bacterial protein Ccl1 is divided into two independently transcribed reading frames. *MGG Mol Genet.* 1996;252(3):292–302. <https://doi.org/10.1007/bf02173775>.
67. Kim B, Kim K, Yang TJ, Kim S. Completion of the mitochondrial genome sequence of onion (*Allium cepa* L.) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the ccmF N gene split. *Curr Genet.* 2016;62(4):873–85. <https://doi.org/10.1007/s00294-016-0595-1>.
68. Choi IS, Ruhlman TA, Jansen RK. Comparative mitogenome analysis of the genus *trifolium* reveals independent gene fission of ccmFn and intracellular gene transfers in Fabaceae. *Int J Mol Sci.* 2020;21(6). <https://doi.org/10.3390/ijms21061959>.
69. Christensen AC. Genes and junk in plant mitochondria-repair mechanisms and selection. *Genome Biol Evol.* 2014;6(6):1448–53. <https://doi.org/10.1093/gbe/evu115>.

70. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*. 2003;424:197. <https://doi.org/10.1038/nature01743>.
71. Rice DW, Alvarson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *amborella*. *Science*. 2013;342(6165):1468.
72. Park S, Ruhlman TA, Weng ML, Hajrah NH, Sabir JSM, Jansen RK. Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of geranium. *Genome Biol Evol*. 2017;9(6):1766–80. <https://doi.org/10.1093/gbe/evx124>.
73. Su L, Fukushima T, Prior A, Baruch M, Zajdel TJ, Ajo-Franklin CM. Modifying cytochrome c maturation can increase the bioelectronic performance of engineered *Escherichia coli*. *ACS Synth Biol*. 2020;9(1):115–24. <https://doi.org/10.1021/acssynbio.9b00379>.
74. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
75. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. <https://doi.org/10.1186/s13059-020-02154-5>.
76. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
77. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
78. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol*. 2019;1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
79. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST plus : architecture and applications. *BMC Bioinformatics*. 2009;10. <https://doi.org/10.1186/1471-2105-10-421>.
80. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. *Science*. 2008;320(5875):486–8. <https://doi.org/10.1126/science.1153917>.
81. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
82. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300>.
83. Al-Mohanna T, Ahsan N, Bokros NT, Dimlioglu G, Reddy KR, Shankle M, et al. Proteomics and proteogenomics analysis of sweetpotato (*Ipomoea batatas*) Leaf and Root. *J Proteome Res*. 2019;18(7):2719–34. <https://doi.org/10.1021/acs.jproteome.8b00943>.
84. Deutsch EW, Mendoza L, Shteynberg D, Slagel J, Sun Z, Moritz RL. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin Appl*. 2015;9(7–8):745–54. <https://doi.org/10.1002/prca.201400164>.
85. Yang J, Moeinzadeh MH, Kuhl H, Helmuth J, Xiao P, Haas S, et al. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat Plants*. 2017;3(9):696–703. <https://doi.org/10.1038/s41477-017-0002-z>.
86. Wu S, Lau KH, Cao Q, Hamilton JP, Sun H, Zhou C, et al. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat Commun*. 2018;9(1):4580. <https://doi.org/10.1038/s41467-018-06983-8>.
87. Sierro N, Batten JN, Ouadi S, Bakaher N, Bovet L, Willig A, et al. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun*. 2014;5:3833. <https://doi.org/10.1038/ncomms4833>.
88. Amborella Genome P. The Amborella genome and the evolution of flowering plants. *Science*. 2013;342(6165):1241089. <https://doi.org/10.1126/science.1241089>.
89. Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat Genet*. 2016;48(6):657–66. <https://doi.org/10.1038/ng.3565>.
90. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 2006;313(5793):1596–604. <https://doi.org/10.1126/science.1128691>.
91. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, et al. The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res*. 2007;35(Database issue):D883–7. <https://doi.org/10.1093/nar/gkl976>.
92. Tomato GC. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012;485(7400):635–41. <https://doi.org/10.1038/nature11119>.
93. Sharma SK, Bolser D, de Boer J, Sonderkaer M, Amoros W, Carboni MF, et al. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3 (Bethesda)*. 2013;3(11):2031–47. <https://doi.org/10.1534/g3.113.007153>.
94. Olsen JL, Rouze P, Verhelst B, Lin YC, Bayer T, Collen J, et al. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature*. 2016;530(7590):331–5. <https://doi.org/10.1038/nature16548>.
95. Fei Z, Joung JG, Tang X, Zheng Y, Huang M, Lee JM, et al. Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics. *Nucleic Acids Res*. 2011;39(Database issue):D1156–63. <https://doi.org/10.1093/nar/gkq991>.
96. Hirsch CD, Hamilton JP, Childs KL, Cepela J, Crisovan E, Vaillancourt B, et al. Spud DB: a resource for mining sequences, genotypes, and phenotypes to accelerate potato breeding. *The Plant Genome*. 2014;7(1). <https://doi.org/10.3835/plantgenome2013.12.0042>.
97. Lim B, Cheng S. Organelle transcriptomes in plants. *Transcriptomics: Open Access*. 2013;02(01). <https://doi.org/10.4172/2329-8936.10006106>.
98. Liu Y, Su W, Wang L, Lei J, Chai S, Zhang W, et al. Integrated transcriptome, small RNA and degradome sequencing approaches proffer insights into chlorogenic acid biosynthesis in leafy sweet potato. *Plos One*. 2021;16(1):e0245266. <https://doi.org/10.1371/journal.pone.0245266>.
99. Organelle genome sequencing of morning glories raw sequence reads. NCBI BioProject PRJNA737311. (2021). <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA737311>
100. Feng Y. Convolvulaceae-mitogenome-project. (2021). GitHub <https://github.com/fengyanlei33/Convolvulaceae-mitogenome-project>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

