Mitochondrial genomes of two parasitic *Cuscuta* species show little evidence of horizontal gene transfer and retain unusually fragmented *ccmFC* genes

Anderson, Benjamin M*[1]; Krause, Kirsten[2]; and Petersen, Gitte*[1]

[1]Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden

[2]Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Breivika, 9037 Tromsø, Norway

*Authors for Correspondence: Benjamin Anderson, banderson2914@gmail.com; Gitte Petersen, gitte.petersen@su.se

Abstract

Background: The intimate association between parasitic plants and their hosts favours the exchange of genetic material, potentially leading to horizontal gene transfer (HGT) between plants. With the recent publication of several parasitic plant nuclear genomes, there has been considerable focus on such non-sexual exchange of genes. To enhance the picture on HGT events in a widely distributed parasitic genus, *Cuscuta* (dodders), we assembled and analyzed the organellar genomes of two recently sequenced species, *C. australis* and *C. campestris*, making this the first account of complete mitochondrial genomes (mitogenomes) for this genus.

Results: The mitogenomes are 265,696 and 275,898 bp in length and contain a typical set of mitochondrial genes, with ten missing or pseudogenized genes often lost from angiosperm

mitogenomes. Each mitogenome also possesses a structurally unusual *ccmFC* gene, which exhibits splitting of one exon and a shift to trans-splicing of its intron. Based on phylogenetic analysis of mitochondrial genes from across angiosperms and similarity-based searches, there is little to no indication of HGT into the *Cuscuta* mitogenomes. A few candidate regions for plastome-to-mitogenome transfer were identified, with one suggestive of possible HGT. **Conclusions:** The lack of HGT is surprising given examples from the nuclear genomes, and may be due in part to the relatively small size of our *Cuscuta* mitogenomes, limiting the capacity to integrate foreign sequences.

Key words: chloroplast, dodder, intracellular transfer, mitogenome.

Background

Horizontal gene transfer (HGT) may be broadly defined as the movement of genetic material between species by a means other than normal reproduction (e.g. fusion of gametes). The process has been inferred within and between both prokaryotes [1] and eukaryotes [2], and is now thought to potentially have a role in adaptation, as shown in recent examples from plants [3, 4]. While the concept is sometimes extended to include transfers between cellular compartments (e.g. chloroplast to mitochondrion), we here refer to HGT as exchange of genetic material between organisms and consider it distinct from intracellular transfers.

Mechanisms that facilitate HGT between plants have been postulated to include direct contact via e.g. wounds/grafting [5], parasitic haustoria [6], epiphytic interactions [7] and illegitimate pollination [8]. Indirect routes have also been suggested, such as via soil, bacteria, viruses or endophytic fungi [8]. Early cases of detected HGT in plants [9, 10] were principally found in their mitochondrial genomes (mitogenomes), possibly partly as a result of the greater sequence availability at the time compared to the nuclear genome. Cellular mechanisms such as active DNA uptake [11] and organelle fusion and fission [12, 13] have

been referenced to explain why HGT is more prevalent in the mitogenome, at least compared with the extensively sequenced chloroplast genome (plastome). While examples of HGT from nuclear genomes are accumulating, mitogenome examples still remain a substantial portion of the evidence for HGT in plants.

A group of plants whose lifestyle makes them more likely to experience HGT are parasites [14, 15], largely because of their close physical interaction with their hosts via specialised connective organs called haustoria. Parasitic plants rely on other plants for water and often for a source of carbon, having some degree of photosynthetic ability (hemiparasites) or having lost it (holoparasites). The altered need for photosynthetic capacity has impacted the size and gene content of plastomes of many parasitic plants [16, 17]. Parasites in the family Viscaceae also have markedly divergent and reduced gene content in their mitogenomes [18–20], but such a reduction was not observed in other groups of parasites [20, 21]. The more common feature associated with parasitic plant mitochondria is the presence of foreign DNA. An extreme example of incorporated foreign DNA in a mitogenome is *Lophophytum mirabile* (Balanophoraceae), which is inferred to have acquired approximately 80% of its genes and 60% of its sequence from its host [22, 23]. Another example includes members of Rafflesiaceae, for which up to 40% of their mitochondrial genes are inferred to have been transferred, with many of the putatively transferred genes shown to be actively transcribed [24].

Parasitic plants in the genus *Cuscuta* are twining stem parasites with reduced/absent leaves that rely on their host plants for both water and a carbon source obtained via haustorial connections [25]. Photosynthetic ability is variably retained in the genus, and there is a diversity of chloroplast structure and plastome gene content [26–30]. The range of possible host species for *Cuscuta* ranges from narrow to especially broad [25], and the haustoria that *Cuscuta* plants use to obtain nutrients from their hosts are also capable of

transmitting genetic material in the form of RNA [6] and DNA [4, 31], making *Cuscuta* a potentially ideal study system for exploring the phenomenon of HGT.

A number of studies have implicated *Cuscuta* species as sources and recipients of HGT events. Studies of host *Plantago* [32, 33] and *Geranium* [34] species show evidence of multiple mitochondrial genes likely obtained from *Cuscuta* parasites. Multiple transcriptomic surveys identified candidate nuclear genes in *Cuscuta* that may have been acquired by HGT [35–37]. Nuclear genome assemblies coupled with transcriptomic work have also revealed numerous candidates for HGT in *Cuscuta campestris* [4, 31]. A recent report of an *Agrobacterium* gene in three *Cuscuta* species could also represent an HGT from another plant, though it could have also been acquired directly from the bacterium [38]. To date the evidence suggests there is active exchange of genetic information between *Cuscuta* and its hosts, with donations of mitochondrial genes and acquisitions of multiple nuclear genes, but not yet any indication of imported mitochondrial genetic material.

While nuclear genomes for *Cuscuta australis* [39] and *C. campestris* [31] have been assembled, our knowledge of the mitogenome in *Cuscuta* is limited to a subset of genes from *Cuscuta gronovii* [34]. Given the lack of knowledge about *Cuscuta* mitogenomes and the expectation that the mitogenome is a prime location to look for HGT, we set out to assemble and characterise mitogenomes for *Cuscuta australis* and *C. campestris*. The complete genomes will enable us to assess the evidence for HGT, and secondarily intracellular transfers between genomic compartments, in *Cuscuta* mitogenomes, and test the unspoken assumption that examples of HGT in *Cuscuta* should be abundant.

Methods

We downloaded from GenBank portions of previously published datasets from whole genome sequencing studies of *Cuscuta australis* and *C. campestris* [31, 39]. For *C. australis*, we downloaded 150 bp paired-end reads from an Illumina 350 bp insert library (SRA

accession SRR5851367; 155 M reads) and PacBio long reads (SRA accessions SRR5851361–6; 711 k reads). For *C. campestris*, we downloaded 301 bp paired-end reads from Illumina 800 bp insert libraries (SRA accessions ERR1917169–71; 9.5 M reads) and PacBio long reads (SRA accessions ERR1910865–7; 490 k reads). Illumina short reads were evaluated with FastQC [40] before and after read deduplication, adapter and low quality base trimming, error correction and read merger using Clumpify, BBDuk, BBMerge and Tadpole in the BBMap package v. 38.08 [41; https://sourceforge.net/projects/bbmap/]. As part of initial assembly, long reads were corrected then trimmed using default settings in Canu v. 1.8 [42]. After initial quality control, the datasets consisted of 1 M merged and 122 M unmerged short reads and 399 k long reads for *C. australis*, and 316 k merged and 7 M unmerged short reads and 79 k long reads for *C. campestris*.

For transcriptome data of *C. australis* [39], we downloaded 150 bp paired-end reads for various tissues (SRR6664647–54; 411 M reads). After read deduplication and adapter and low quality base trimming using Clumpify and BBDuk, 268 M reads were retained with an average read length of 142 bp (72% of reads were untrimmed).

Organelle genome assemblies

We approached organellar assembly by running separate short and long read *de novo* assemblies, searching resulting contigs using BLAST v. 2.6.0+ [43–45] for hits to reference coding sequences (CDS), then using those contigs with hits as targets for further mapping with BBMap and assembly.

Short read datasets were first filtered using the Kmer Analysis Toolkit (KAT) v. 2.4.2 [46] to select a high coverage subset of reads for easier assembly. Coverage targets were based on mapping reads to sequences from reference plastome (*Cuscuta exaltata*, NC_009963) and mitogenome (*Ipomoea nil*, NC_031158) CDS. Based on the mapping, we chose kmer depth filters of 500–25000 selecting ~35 M reads and 30–2100 selecting ~700 k

reads for *C. australis* and *C. campestris*, respectively. The filtered read sets were assembled using Unicycler v. 0.4.8 [47], which when given only short reads optimizes a SPAdes v. 3.14.0 [48] assembly.

Long read datasets were assembled in Canu v. 1.8 [42] using default settings and with genome size estimates of 280 M and 550 M for *C. australis* and *C. campestris*, respectively. In order to complete one run, we set the stopOnLowCoverage parameter to 0.1 for *C. campestris*.

We examined contigs from the short and long read assemblies in Bandage v. 0.8.1 [49] to identify those with BLAST hits to the reference CDS sequences (*Cuscuta exaltata* and *Ipomoea nil*) and which had similar depths (for distinguishing plastid and mitochondrial contigs). Contigs with hits and identified as putatively plastid or mitochondrial were used as targets for mapping the full sets of short reads to create input datasets for Unicycler. The mapped reads were assembled along with long reads (corrected and trimmed with Canu) in Unicycler, which assembles long reads with miniasm [50] and Racon v. 1.4.3 [51] and uses the assembly to resolve repeats and create bridges in the SPAdes small read assembly graph.

Plastome assemblies were obtained after a single round of Unicycler, producing single circular contigs. In order to complement our assembly approach, we also used NOVOPlasty v. 3.7.2 [52] with the deduplicated and adapter-trimmed (but not error corrected or merged) short read datasets. We used an *rbcL* sequence from *Cuscuta exaltata* as a seed, which NOVOPlasty uses to select an initial matching read and begin iterative extension until it can circularize an assembly. The resulting two options per assembly (different configurations across the inverted repeats) were then compared with BLAST to the assemblies from Unicycler to verify the sequences were identical.

The mitogenome assemblies were more challenging, given the presence of long repeats and linear and circular contigs that in some cases could be joined at multiple

150 locations. For *C. australis*, a second round of Unicycler was run after extending contigs by up
151 to 100 kbp using Tadpole (part of the BBMap package) and then mapping short reads again.
152 Simplifying the assembly graph included removing small contigs that were fully present in a
153 larger contig or were too short (e.g. <100 bp). We then checked long read contigs from the
154 initial assembly in Bandage to ensure they were covered by hits from the remaining
155 assembly. Since there were four linear and four circular contigs, we did not attempt to merge
156 them into a single contig. For *C. campestris*, the first round of Unicycler produced four circles
157 along with three small linear contigs (totalling <5 kb), which were either not found in the long
158 read contigs or had hits to the plastome that were already present in one of the circles. Since
159 the four circles had shared long repeats, these were examined using long reads to determine
160 if multiple configurations were present, suggesting the separate contigs are sometimes
161 contiguous. Long reads were mapped and aligned to extracted repeats with Minimap2 v.
162 2.14-r883 [50], Samtools v. 1.7 [53] and MAFFT v. 7.310 [54] and using a custom Python
163 script including 'pysam' (https://github.com/pysam-developers/pysam). When the reads
164 indicated multiple configurations across the repeat, we manually joined the circles. The
165 circles all shared repeats that showed multiple configurations, so we were able to merge the
166 four circles into a single circular contig.

167 **Read mapping and annotation**

168 In order to check the mitogenome assemblies for clear errors, we conservatively
169 mapped short and long reads then manually compared the resulting BAM files in IGV v. 2.6+
170 [55]. Short reads were mapped using BBMap with settings pairedonly=t minid=0.9
171 ambiguous=random maxindel=100 strictmaxindel=t pairlen=1000. Long reads were mapped
172 using Minimap2 in mode 'map-pb' and filtered with Samtools with option -F 2308, followed by
173 filtering the resulting BAM file with a custom Python script using 'pysam' to only retain
174 mappings where reads covered the target for at least 80% of their length. For *C. australis*,

175 only one location suggested a small (28 bp) sequence that needed to be deleted (absent in
176 most long reads). For *C. campestris*, one location of a plastome insert needed three single
177 base pair corrections and a small (9 bp) deletion to match the long reads that spanned the
178 region rather than the higher depth plastid reads.

179 Plastomes and mitogenomes were annotated using custom scripts to BLAST
180 (BLASTN, BLASTX, default search settings) against nucleotide and protein databases built
181 using reference plastid (14 from 13 genera and 13 families, see Table S1, Additional file 1)
182 and mitochondrial (172 from 130 genera and 62 families, see Table S2, Additional file 1)
183 genomes, an approach similar to that used in Mitofy [56]. For mitogenomes, we also used
184 the references to build protein alignments for constructing a HMMER v. 3.3 [57;
185 http://hmmer.org] database that we used to search open reading frames (ORFs) found using
186 ORFfinder v. 0.4.3 (NCBI; https://www.ncbi.nlm.nih.gov/orffinder/). Following searches, hits
187 were examined and adjusted manually, with input from MFAnnot v. 1.35
188 (https://github.com/BFL-lab/Mfannot) to select intron boundaries. Transfer RNAs were also
189 detected and checked with tRNAscan-SE 2.0 [58]. Pseudogenes were subjectively annotated
190 if hits had clear frameshifts and/or internal stop codons, or if they were only fragments of full
191 length genes (potentially broken due to alternative paths through the assembly).

192 Dispersed repeats were detected by searching the organellar genomes against
193 themselves using BLASTN with an E value of 1e-10 and filtering out hits less than 50 bp and
194 lower than 90% identity. Overlapping hits were then filtered with a custom python script, and
195 the percentage repeat content calculated. For visualization, the genomes were searched
196 against themselves using Circoletto [59] with an E value chosen to approximate a 90%
197 identity threshold (1e-70 and 1e-40 for *C. australis* and *C. campestris*, respectively). As a
198 proxy for recombinational activity, we used long repeats (>500 bp, >90% identity) and
199 mapped long reads. Up to 50 long reads that mapped to each repeat were aligned with the
200 repeat and evaluated manually for whether they showed indications of incongruence on

either side of the repeat. A proportion of the reads showing incongruence was used to assess recombinational activity. Tandem repeat content was determined using Tandem Repeat Finder v. 4.09 [60] with recommended settings: 2 7 7 80 10 50 500. Organelle genome figures were based on output from OGDRAW [61].

To investigate the transcription and splicing of the mitochondrial *ccmFC* gene, we mapped mRNA reads using BBMap with the setting pairedonly=t against three DNA sequences manually stitched together and separated by N's from the annotated *ccmFC* locations in the *C. australis* assembly. We also aligned *ccmFC* DNA sequences from both *Cuscuta* species with *ccmFC* genes and CDS from reference mitogenomes using MAFFT to determine the locations of the exon and intron boundaries. We visualized alignments and mapping in Jalview v. 2.11.0 [62] and IGV v. 2.8.0 [55]. Intron domains were roughly determined by iteratively running the Mfold web server [63] using the ccmFci829 from *Nicotiana tabacum* (NC_006581). RNA editing sites were predicted with PREP-MT [64] to compare with those observed in the transcriptomic reads.

To evaluate the notable gene structure of the final portion of *nad1*, we aligned sequences from *Ipomoea nil*, *C. australis* and *C. campestris* using Mauve v. snapshot_2015-02-13 [65].

Potential intracellular transfers

To detect the presence of transfers from a plastome to the mitogenome, we first searched each *Cuscuta* plastome against its corresponding mitogenome using BLASTN with default parameters. Next, we searched each mitogenome against the plastome of *Ipomoea nil* (NC_031159) using BLASTN with default parameters, to include a close relative with a more complete set of plastid genes. We also searched the mitochondrial ORFs against a HMMER database built with the reference plastomes to detect whether there were intact

225 portions of plastid genes present and to interpret blast hits. Broader searches used the NCBI
226 non-redundant nucleotide database (https://blast.ncbi.nlm.nih.gov/).
227 To determine whether any missing mitochondrial genes were present in the nuclear
228 genome, and whether there were other transfers between the mitogenome and the nuclear
229 genome, we conducted BLAST searches of our assemblies and of our reference
230 mitogenomes against the two published nuclear genomes (*C. australis*: GCA_003260385.1,
231 *C. campestris*: PRJEB19879). For assessment of total hit coverage of our assemblies, we
232 ignored hits <250 bp and with <80% identity. For searches for missing genes, we ignored hits
233 with <50% identity.

234 **Horizontal gene transfer**

235 We investigated HGT using a phylogenetic approach, in which we built phylogenetic
236 trees for each mitochondrial gene present in *Cuscuta* (as well as one based on all genes
237 concatenated) and examined whether *Cuscuta* sequences were supported as grouping with
238 their close relatives (particularly *Ipomoea nil*) or with more distantly related plants. Using
239 reference mitogenomes (see Table S2, Additional file 1) and our *Cuscuta* assemblies along
240 with data from *C. gronovii* (GenBank accessions KP940494–514), we extracted all CDS
241 protein translations and corresponding nucleotide sequences, as well as nucleotide
242 sequences for ribosomal RNAs. We removed some sequences in cases of high divergence
243 (e.g. for *Viscum scurruloideum*) and when alignments suggested problematic annotations,
244 and in some cases we had to reverse complement ribosomal RNA sequences. For each
245 gene, we created protein alignments with ClustalOmega v. 1.2.4 [66] using default settings.
246 We then converted the protein alignments back to nucleotide alignments with PAL2NAL v. 14
247 [67] using default settings. For ribosomal RNA sequences, we aligned with MAFFT v. 7.310
248 using options --localpair and --maxiterate 1000. For each alignment we used a custom
249 python script to eliminate positions with >50% missing data, then ran RAxML v. 8.2.12 [68]

250 with rapid bootstrapping and a search for best tree in a single run, opting for 1000 bootstrap
251 replicates under the GTRGAMMA model. Finally, we plotted the resulting trees using R
252 package ape [69], rooting them on non-angiosperm seed plants, and manually examined
253 whether *Cuscuta* sequences grouped together, with their close relative *Ipomoea nil*, and
254 among lamiids or whether they grouped elsewhere with bootstrap support >80%.

255 To explore intergenic sequences, we searched our reference mitogenomes against
256 our *Cuscuta* assemblies using BLASTN with default parameters, removing hits <250 bp and
257 with <80% identity. Using custom scripts we filtered blast hits so that only the top identity hit
258 and those within 2% identity were retained for a given position along the assembly. We
259 plotted hits using the R package Sushi v. 1.26.0 [70]. We then manually examined hits in
260 intergenic regions, looking for where the top hits were not from Solanales and evaluating
261 whether the hits were suggestive of longer, potentially transferred tracts.

262 In cases of possibly chimeric gene sequences based on top blast hits, we assessed
263 evidence for conversion using GENECONV v. 1.81 [71].

264 Results

265 Organelle genomes

266 The plastomes were assembled as single circular molecules having typical
267 quadripartite structure with two inverted repeats separating small and large single copy
268 regions (see Figs. S1 and S2, Additional file 2). The plastomes are 85,263 and 86,749 bp,
269 with GC contents of 37.8 and 37.7% for *C. australis* and *C. campestris*, respectively. Beyond
270 the inverted repeats (14,070 and 14,348 bp in *C. australis* and *C. campestris*, respectively),
271 there are no dispersed repeats longer than 50 bp and with identity >90%. Tandem repeats
272 make up approximately 0.2 and 0.4% of the plastomes of *C. australis* and *C. campestris*,
273 respectively. Similar to other reduced plastomes of *Cuscuta* species such as *C. gronovii* [28]

and *C. obtusiflora* [27], the plastomes are missing all *ndh* genes (a pseudogenized copy of *ndhB* is present in *C. campestris*), *infA*, *matK*, *psaI*, all *rpo* genes, *rpl23*, *rpl32*, and *rps16*. Missing tRNAs include *trnA*-UGC, *trnG*-UCC, *trnI*-GAU, *trnK*-UUU, *trnR*-ACG, and *trnV*-UAC.

The mitogenomes were assembled as 265,696 and 275,898 bp, with GC contents of 44.0 and 44.3% for *C. australis* and *C. campestris*, respectively (Fig. 1). Dispersed repeats (see Fig. S3, Additional file 2) comprise approximately 21.8 and 34.1% while tandem repeats make up approximately 5.0 and 2.5% of the mitogenomes of *C. australis* and *C. campestris*, respectively. Recombinational activity as assessed with long repeats varied from 10–48% (average = 29%, n=6) and 33–47% (average = 41%, n=10) for *C. australis* and *C. campestris*, respectively. For each mitogenome, we annotated 31 protein-coding genes, 3 ribosomal rRNA genes, and 19 tRNAs (some in multiple copies; Table 1). Compared to *C. australis*, *C. campestris* has an extra *trnQ*-UUG of plastome origin (within a possibly transferred region), but is missing *trnT*-UGU. Common missing or pseudogenized genes for both genomes include *sdh3* and *sdh4*, *rpl2* and *rpl10*, and six *rps* genes (*rps2, 7, 10, 11, 14, 19*). Some extra fragments of already present genes broken up by possible alternative paths through the assembly include *atp1* and *cox1* for both species, and *atp6*, *rps12* and *rps3* for *C. campestris*.

Table 1. Gene presence in the mitogenomes of *Cuscuta australis*, *C. campestris* and their close relative *Ipomoea nil*.

Gene	*C. australis*	*C. campestris*	*Ipomoea nil*	Gene	*C. australis*	*C. campestris*	*Ipomoea nil*
atp1	Y	Y	Y	*rps10*	-	-	Y
atp4	Y	Y	Y	*rps11*	-	-	-
atp6	Y	Y	Y	*rps12*	Y	Y	Y
atp8	Y	Y	Y	*rps13*	Y	Y	Y
atp9	Y	Y	Y	*rps14*	pseudo	pseudo	Y
ccmB	Y	Y	Y	*rps19*	pseudo	pseudo	Y
ccmC	Y	Y	Y	*sdh3*	-	-	-
ccmFC	Y	Y	pseudo	*sdh4*	pseudo	pseudo	Y
ccmFN	Y	Y	Y	*rrn5*	Y	Y	Y
cob	Y	Y	Y	*rrn18*	Y	Y	Y
cox1	Y	Y	Y	*rrn26*	Y	Y	Y
cox2	Y	Y	Y	*trnC*-GCA	Y	Y	Y
cox3	Y	Y	Y	*trnD*-GUC	Y	Y	Y
matR	Y	Y	Y	*trnE*-UUC	Y	Y	Y
mttB	Y	Y	pseudo	*trnF*-GAA	Y	Y	Y
nad1	Y	Y	Y	*trnG*-GCC	Y	Y	Y
nad2	Y	Y	Y	*trnH*-GUG	Y	Y	Y
nad3	Y	Y	Y	*trnI*-CAU	Y	Y	Y
nad4	Y	Y	Y	*trnK*-UUU	Y	Y	Y
nad4L	Y	Y	Y	*trnM*-CAU	Y	Y	Y
nad5	Y	Y	Y	*trnM*-CAU-cp	Y	Y	-
nad6	Y	Y	Y	*trnfM*-CAU	Y	Y	Y
nad7	Y	Y	Y	*trnN*-GUU	Y	Y	Y
nad9	Y	Y	Y	*trnP*-UGG	Y	Y	Y
rpl2	-	-	-	*trnQ*-UUG	Y	Y	Y
rpl5	Y	Y	Y	*trnQ*-UUG-cp	-	Y	-
rpl10	-	-	Y	*trnS*-GGA	-	-	Y
rpl16	Y	Y	Y	*trnS*-GCU	Y	Y	Y
rps1	Y	Y	Y	*trnS*-UGA	Y	Y	Y
rps2	-	-	-	*trnT*-UGU	Y	-	Y
rps3	Y	Y	Y	*trnW*-CCA	Y	Y	Y
rps4	Y	Y	Y	*trnY*-GUA	Y	Y	Y
rps7	-	-	-				

Table showing gene presence, where present genes are indicated with a 'Y', while absent genes are shown with a dash. Pseudogenes are indicated with 'pseudo'.

297 The *ccmFC* gene was problematic to annotate in both species (annotated as a
298 pseudogene in *C. gronovii* and *Ipomoea nil*) and may not be properly transcribed, as it
299 appears to be both trans-spliced and missing a portion of the first exon. Based on alignments
300 with other angiosperms, including the inton-less *Platycodon grandiflorus* (NC_035958), it
301 appears that the *ccmFC* gene is split across three locations in both of the *Cuscuta*
302 mitogenomes (Fig. 2a). A large portion of exon 1 is separated from the 3' end of exon 1 and
303 most of the group II intron. The remainder of the intron plus exon 2 are in a third location.
304 Mapping mRNA reads resulted in low coverage of the *ccmFC* gene, but provided some
305 support (four read pairs) for successful trans-splicing of the 3' end of exon 1 and exon 2 (see
306 Fig. 2a). There was no indication of an mRNA product that included both portions of exon 1,
307 however, making it unclear whether the majority of exon 1 is actually translated into part of a
308 functional product. PREP-MT predicted 14 editing sites, three of which didn't have
309 transcriptome support, and missed four sites that appeared to be RNA edited based on
310 transcriptome read mapping (see Fig. S4, Additional file 2).
311 Another odd gene structure found in both *Cuscuta* species is the final portion of *nad1*,
312 namely exon 4, *matR* and exon 5 (see Fig. 2b). In our two *Cuscuta* assemblies, exon 5 is
313 found in the reverse orientation on the opposite strand, while exon 4 is located on the same
314 strand as *matR*. This may indicate nad1i728 has shifted to trans-splicing (as seen in some
315 other angiosperms, e.g. *Nicotiana tabacum*, NC_006581.1), but if that is the case, it seems
316 odd that exon 5 is in such close proximity to the rest of the intron and exon 4.

317 **Potential intracellular transfers**

318 Using the respective plastome of each *Cuscuta* species and that of *Ipomoea nil* as
319 BLAST queries, we plotted plastid hits on the mitogenome assemblies (see Fig. 1). Excluding
320 hits that occurred in the ribosomal genes or *atp1*, which are similar to plastid ribosomal

genes and *atpA*, we found 6 regions with hits (3 in each species), representing approximately 0.4 and 0.9% of the mitogenomes of *C. australis* and *C. campestris*, respectively (Table 2).

Table 2. Results of BLASTN searches of *Cuscuta* mitogenomes using queries of *Cuscuta* and *Ipomoea* plastomes.

Species	**Hit length (bp)**	**Hit coords (contig)**	**% ID**	**HMMER hit**	**HMMER coords (contig)**	**NCBI megablast**
C. australis	288 138	17586..17873 (2) 17432..17569 (2)	64 93	*psbB*	17868..17641 (2) 17580..17374 (2)	*C. gronovii* (mitogenome); Malpighiales Myrtales; Cucurbitales; Malpighiales
C. australis	255 210	15995..16249 (2) 15698..15907 (2)	92 72	*ndhB*	16251..15917 (2)	Geraniales; Gentianales; others Celastrales; Rosales; others
C. australis	146	84001..84146 (1)	76	-	-	*Ipomoea* spp.
C. campestris	120 76 138	143582..143701 143723..143798 143823..143960	72 67 93	*psbB*	143587..143760 143812..143970	*C. gronovii* (mitogenome); Malpighiales *C. gronovii* (mitogenome) Cucurbitales; Malpighiales; Myrtales
C. campestris	343 219	144976..145318 145328..145546	80 75	*ndhB*	144886..144963 144974..145377	Gentianales; Geraniales; others Rosales; Celastrales; others
C. campestris	1616	61212..62827	93–95	*psbA*	61904..62623 62636..62826	*Cuscuta* spp.

Table showing locations of hits in the BLAST and HMMER searches. NCBI megablast results are indicated for top scoring results, sometimes by order for simplicity, with multiple results shown when top hits were not clearly better.

Both species share two regions corresponding to HMMER hits to plastid genes *psbB* and *ndhB*. Examining the regions using megablast searches of the NCBI non-redundant nucleotide database indicated top hits to orders outside Solanales. The *psbB* hit region was also detected in *C. gronovii* mitogenome sequences (KP940494). In the case of the *ndhB* hits, multiple orders were represented in the top hits, making it difficult to determine the best

335 candidate order for a possible foreign transfer. In both *Cuscuta* species, the *psbB* and *ndhB*
336 hit regions are within 2 kb of each other, possibly indicating a larger transferred region.
337 The third region found in *C. australis* is a hit to the *Ipomoea* plastome with no clear
338 HMMER hits. Searching NCBI also gave top hits to *Ipomoea*. The third region found in *C.*
339 *campestris* corresponds to HMMER *psbA* hits, though the region has a frameshift and does
340 not represent a full-length gene ORF. Searches of NCBI gave top hits to *Cuscuta* plastomes.
341 The region also includes a *trnQ*-UUG of plastid origin, which is co-located near *psbA* in the
342 *C. campestris* plastome.
343 Searches of the nuclear genomes produced markedly different results for the two
344 *Cuscuta* species. In *C. australis*, the ~260 Mbp assembly hit our mitochondrial assembly for
345 cumulatively ~27% of its length, but in *C. campestris*, the ~477 Mbp assembly hit our
346 mitochondrial assembly for ~99.8% of its length or essentially the entire assembly. Four of
347 the scaffolds in the *C. campestris* assembly are essentially entirely mitochondrial, and four
348 more are > 70% mitochondrial. These scaffolds could tentatively indicate recent large-scale
349 integration of mitochondrial sequences into the nuclear genome, but could also be simple
350 assembly artifacts since mitochondrial contaminations are difficult to identify.
351 Potential candidates for missing mitochondrial genes were identified in both nuclear
352 genomes for *rps2*, *rps10*, *rps11*, *rps14*, *rps19*, *sdh3* and *sdh4*. All the candidates had intact
353 open reading frames except for *sdh3*, which had no determined start codon. BLAST
354 searches of NCBI mostly gave top hits to the same mitochondrial genes (or a nuclear copy),
355 and mostly to *Ipomoea* species, though *rps10* and *rps19* had top hits to Amaranthaceae, and
356 *sdh4* to Fabaceae. Phylogenetic analysis of the candidates with copies from our reference
357 mitogenomes consistently put the nuclear copies on long branches and with low support for
358 group placement (see Fig. S5, Additional file 3), making the origin of the sequences unclear.
359 Searching for indications of mitochondrial targeting sequences with Predotar v. 1.04 [72]

360 indicated possible mitochondrial targeting (p 0.21–0.44) for *rps10*, *rps14*, *rps19* and *sdh3* (if
361 an 'M' is added in place of the undetermined start codon).

362 Horizontal gene transfer

363 Phylogenetic analysis of the concatenation of all mitochondrial genes strongly
364 supports the grouping of *Cuscuta* within Solanales (Fig. 3). Based on the individual gene
365 trees with sampling across seed plants, the majority of genes (26/34, 76%) in the three
366 *Cuscuta* species are strongly supported (>80% bootstrap support) as grouping together and
367 sister to their close autotrophic relative *Ipomoea nil*, a pattern consistent with vertical rather
368 than horizontal inheritance (see Fig. S6, Additional file 4). Exceptions to this pattern are *atp9*
369 (64% sister to *Ipomoea*), *cox2* (74%), *nad3* (unsupported), *nad4L* (not sister to *Ipomoea*, but
370 88% support together with *Ipomoea* in Solanales), *rpl16* (74%), *rps1* (76%), *rps12*
371 (unsupported), and *rrn5* (unresolved) (note that *ccmFC* and *rrn5* from *C. gronovii* were not
372 included). In the cases with no support for a sister relationship with *Ipomoea*, there is no
373 strong support for a grouping elsewhere, as would be expected for HGT.
374 Intergenic sequences generally lacked clear evidence of longer (>500 bp) contiguous
375 groups of hits to single orders outside Solanales (see Figs. S7 and S8, Additional file 5).
376 Some exceptions could be identified, including hits to Asterales, Caryophyllales, Fabales,
377 Gentianales, Lamiales and Vitales, totalling approximately 5.9 and 3.9% of the mitogenomes
378 of *C. australis* and *C. campestris*, respectively. Top hits (by 2% identity) covering *cox3*, *nad4L*
379 and *rps13* were not to Solanales, though phylogenetic evidence indicates *cox3* and *nad4L*
380 group there, and *rps13* is associated with *Ipomoea nil*.
381 In portions of some genes there were indications that blast hits to other orders were
382 higher than to Solanales (*cox2* exon 1, *nad4* exon 3, and *nad5* exon 5), so we ran
383 GENECONV to detect possible conversion. In two cases there were some sequences
384 showing significant ($P < 0.05$) signal of conversion (4 sequences for *cox2* and 2 for *nad5*),

385 but not involving the *Cuscuta* sequences. For *nad4*, there were many significant
386 comparisons, but all but one involved *Vicia faba* (KC189947), suggesting it may have
387 undergone gene conversion, consistent with its long branch in the phylogenetic tree. There
388 was no significant indication that the *Cuscuta* sequences were chimeric.

389 Discussion

390 Recently sequenced genomes of parasitic *Cuscuta* species revealed extensive HGT
391 events into their nuclear genomes from a broad range of non-parasitic angiosperms that
392 serve as their hosts. It was a reasonable expectation that there would be high transfer
393 activity of genetic material into the mitogenomes of these parasites as well. The organellar
394 genomes of the two *Cuscuta* species that we have assembled and annotated here, however,
395 are not clearly different in gene content from other *Cuscuta* plastomes or typical angiosperm
396 mitogenomes. Based on phylogenetic analysis, we found little evidence for HGT into the
397 mitogenomes of the two species, an unexpected result given the parasitic lifestyle of *Cuscuta*
398 and evidence for HGT candidates in their nuclear genomes. In the process of examining the
399 mitogenomes, we also identified a novel structure of the *ccmFC* gene in *Ipomoea* and
400 *Cuscuta*, indicative of a shift to trans-splicing of the group II intron and possibly a shorter
401 product.

402 Gene losses and possible intracellular transfers

403 All of the ten genes missing from the two *Cuscuta* mitogenomes are commonly found
404 to be lost during angiosperm evolution [73, 74]. That *rps2* and *rps11* are missing is consistent
405 with the inference of their ancient losses in angiosperms [73]. For seven of the ten missing or
406 pseudogenized genes, potential candidates for nuclear copies were found. Transfers of
407 mitochondrial genes to the nucleus have been inferred to be numerous across plant
408 evolution and seem to be the primary mode of functional gene loss from the mitogenome

409 [75]. Whether the gene products from our nuclear candidates are functional or imported into
410 the mitochondrion is uncertain, but tests for targeting sequences suggested some of them
411 might be targeted to the mitochondrion. The nuclear candidates are on longer branches in
412 phylogenetic trees, suggesting higher substitution rates. This is consistent with observations
413 of higher substitution rates in the nuclear genome compared to the mitogenome [76], and of
414 higher rates in nuclear copies of genes transferred from the mitogenome compared to the
415 mitogenome copies [77].

416 The three missing mitochondrial genes not detected in the nuclear genomes are *rpl2*,
417 *rpl10* and *rps7*. Losses of *rpl2* and *rps7* are frequent among angiosperms, but only for *rpl2*
418 are losses commonly associated with functional transfers to the nucleus [34, 73, 78, 79].
419 Losses of *rpl10* are less common and have been associated with functional replacement by a
420 duplicated nuclear gene normally encoding chloroplast RPL10 in some monocot and
421 Brassicaceae lineages [74, 80, 81]. Searching ORFs from both nuclear genomes for possible
422 copies of RPL10 suggested the presence of one (*C. australis*) or two (*C. campestris*) genes
423 more similar to one of the copies from *Arabidopsis* (NP_196855) than to the other
424 (NP_187843). Our search suggests there is only the chloroplast RPL10 present and not a
425 duplicate targeted to the mitochondrion.

426 Plastid transfers appear to be limited in the mitogenomes, with only a single 1.6 kb hit
427 in *C. campestris* having high similarity to its plastome. The other main candidate regions
428 included hits to *psbB* and *ndhB*, but based on BLAST searches appear to be more similar to
429 regions in distantly related plants. The BLAST results are inconclusive as to the origin of the
430 region, but it could represent a horizontal transfer, possibly via an intracellular plastome
431 transfer to the donor's mitogenome and then a plant to plant transfer between mitogenomes,
432 a common route for plastome to mitogenome HGT [82]. The close association of the *psbB*
433 and *ndhB* hits suggest transfer of a larger contiguous block of DNA. Given *ndhB* is lacking
434 from our *Cuscuta* plastomes (a pseudogene fragment in *C. campestris*), and given the two

genes are not closely associated in the plastomes of *C. campestris* nor *Ipomoea*, it is unlikely the fragment is derived from an ancient *Cuscuta* intracellular plastome-to-mitogenome transfer. If it is instead HGT, then it may have come from a species with close association of those two genes in its plastome, or by two separate transfers from a plastome to close proximity in a mitogenome, then a single transfer from that mitogenome to *Cuscuta*.

Lack of horizontal gene transfer

Phylogenetic evidence for HGT in our *Cuscuta* mitogenome assemblies is lacking, with only a few genes having <80% bootstrap support for a close relationship to *Ipomoea nil*, and those exceptions also with negligible support for alternative relationships. BLAST-based assessment of intergenic sequences did uncover a few longer candidates for possible HGT, but it is unclear how reliable the similarity search is for uncovering the origin of those sequences. Given the dynamic nature of mitogenome size and structure in angiosperms [83–86], our limited sampling of lineages may not include more similar sequences from more closely related plants. The HGT candidates also lack embedded genes for the most part, making it difficult to determine homology and to further support an origin. In cases where BLAST hits to another order cover genes (*cox3*, *nad4L*, *rps13*), phylogenetic analysis of the genes generally contradicts the BLAST-based assignment. The aforementioned presence of sequences of chloroplast origin similar to regions in distantly related plants may also indicate that some HGT has occurred, possibly earlier in the evolution of *Cuscuta* given the region is shared by all the *Cuscuta* sequences we examined.

Why the mitogenomes lack clear evidence of HGT is unclear, given it is prevalent in the nuclear genome of *Cuscuta*. One possible explanation may be related to the size of the mitogenomes we have assembled, which are on the low side for angiosperms. Based on our reference set of mitogenomes (with 168 angiosperms), the average size of the angiosperm mitogenomes is roughly 670 kb, while our assemblies are less than half that. Given there has

been a correspondence observed between mitogenome size and the amount of incorporated plastid transfers [82] and that plant mitochondrial genomes appear to have relaxed restriction on the size of intergenic spacers [e.g. 87], it may be that larger mitogenomes with longer intergenic regions allow for the incorporation of more foreign DNA. Conversely, the smaller mitogenomes in *Cuscuta* may experience more restriction on what can be successfully incorporated.

It is unclear why there has not been an extensive expansion of intergenic material in *Cuscuta* or *Ipomoea* for that matter (265,768 bp). The wide range of sizes of angiosperm mitogenomes are difficult to explain, but recombination rates have been put forward as one potential factor, albeit a complex one [86]. In a study of the massive differences in size amongst *Silene* mitogenomes, Sloan et al. [86] found that the larger mitogenomes experienced lower recombinational activity (<20%) and had more divergent repeats, suggesting reduced gene conversion. Our assembled mitogenomes had relatively high recombinational activity across longer repeat sequences as assessed with long reads (often >20% and >40% for *C. australis* and *C. campestris*, respectively), though with small sample sizes. The difficulty in assembling the mitogenomes and the multiple potential paths through the assemblies are also consistent with multiple conformations across highly similar or identical repeats, possibly indicating ongoing recombination and conversion.

Curious structure of *ccmFC* and its trans-splicing intron

Both *Cuscuta* species and their close autotrophic relative *Ipomoea nil* appear to share a different structure of their *ccmFC* genes than other plants. In all three cases, the first exon is split into two pieces, and the same has happened for the group II intron, making trans-splicing of the intron necessary in order to join the now remotely located parts of the gene. In angiosperm mitogenomes, most introns are group II introns (having six domains) and require multiple protein factors, including some encoded in the nuclear genome, for

485 proper splicing [88]. While the exact mechanism of trans-splicing for group II introns and the
486 splicing factors required remain unclear, this type of splicing has been observed in a number
487 of other angiosperm mitochondrial genes, namely *nad1*, *nad2* and *nad5* [89] and *cox2* [90].
488 Typically, the break point for the two pieces of the intron that are trans-spliced is in domain IV
489 [89, 91], which is where the intron is broken in *Cuscuta* and *Ipomoea*. The evidence
490 presented here, namely the mRNA reads that indicate a product joined from two disparate
491 locations in the genome and the break in domain IV, suggests this is the fifth such example of
492 a gene with a group II intron that has shifted to trans-splicing.

493 Angiosperm *ccmFC* genes examined so far typically have intact rather than broken
494 group II introns. For splicing of the intact intron, an essential nuclear-encoded splicing factor,
495 WTF9, has been identified which shares an RNA-binding domain with another factor involved
496 in intron splicing in plastids [92]. In examining the role of WTF9, des Francs-Small et al. [92]
497 also elucidated the role of *ccmFC* in cytochrome biogenesis, showing that proper splicing of
498 the intron for the formation of a complete protein was essential to the generation of
499 cytochromes c and c_1 in *Arabidopsis*. In addition, splicing-impaired mutants had higher levels
500 of alternative oxidase, suggesting problems with the respiratory chain [92]. It appears that
501 being able to produce a fully intact *ccmFC* product is important for respiratory function, which
502 makes the break in the first exon in *Cuscuta* and *Ipomoea* puzzling. While there is evidence
503 that the intron is trans-spliced and connects the 3' portion of exon 1 with exon 2, there is no
504 indication that the larger 5' portion of exon 1 is connected for translation into the final product.
505 Transcriptomic reads show RNA editing in all parts of the gene (see Fig. 2a and Fig. S4,
506 Additional file 2), similar to predictions from PREP-MT, but whether this indicates an intact
507 product is unclear. Normally, the full *ccmFC* product is thought to be integrated into a larger
508 complex potentially involved in heme ligation [93]. The functionality of at least part of the
509 *ccmFC* gene would be consistent with its conservation in both *Ipomoea* and *Cuscuta*, but the
510 longer branch lengths relative to other members of Solanales (see Fig. S6b, Additional file 4)

511 may suggest some relaxation of selection. Gene loci coding for homologs of the WTF9
512 splicing factor are present in *C. campestris* and *C. australis* (data not shown), excluding this
513 as a reason for the switch. How *Cuscuta* and *Ipomoea* are able to function effectively with
514 their *ccmFC* structure is unclear, and additional work would be helpful to characterise and
515 evaluate the *ccmFC* protein products in this group of plants.

516 While our assembly approach is not guaranteed to have captured the entirety of the
517 mitogenomes (for instance, our *Cuscuta australis* assembly consists of multiple contigs, and
518 our *C. campestris* assembly was circularized from multiple smaller circular contigs), our
519 search approach makes it unlikely that any missing sequences would carry substantially
520 different phylogenetic signals of HGT than the bulk of our assemblies or that they would
521 harbour intact foreign genes or a contiguous copy of *ccmFC*. Alternative structural
522 arrangement of our contigs and construction of our assemblies are also unlikely to affect our
523 conclusions on HGT, but could have some impact on how we have interpreted the structure
524 of *ccmFC*. The lack of mRNA support for a fully intact *ccmFC* product, however, supports our
525 assembly and interpretation of a fragmented gene structure.

526 Conclusions

527 Our search of the assembled mitogenomes of *Cuscuta australis* and *C. campestris*
528 revealed little to no indication of HGT, contrary to our expectations based on examples from
529 their nuclear genomes. This finding suggests that elevated rates of HGT into mitogenomes,
530 where the bulk of examples from parasitic plants have been found to date, is not always the
531 case. The observed amount of HGT appears to also be influenced by species-specific factors
532 such as mitogenome size and recombination frequency that allow for any DNA uptake to be
533 retained.

Additional files

Additional file 1: supplementary Tables S1 and S2. (pdf)

Additional file 2: supplementary Figs. S1–S4. (pdf)

Additional file 3: supplementary Fig. S5. (pdf)

Additional file 4: supplementary Fig. S6. (pdf)

Additional file 5: supplementary Figs. S7 and S8. (pdf)

Additional file 6: fasta alignment files for phylogenetic analyses. (zip)

Abbreviations

CDS: coding sequence; HGT: horizontal gene transfer; ORF: open reading frame.

Availability of data and materials

Organelle genomes have been submitted to GenBank and are awaiting accession numbers.

Sequence alignments used in phylogenetic analysis are included as Additional file 6.

Competing interests

The authors declare that they have no competing interests.

Funding

B. Anderson and G. Petersen were supported by the Department of Ecology, Environment and Plant Sciences, Stockholm University. Some assemblies and analyses were performed on computing resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational

Science (UPPMAX) under Project SNIC 2018/8-336. K. Krause was supported by a grant from the Tromsø Research Foundation [grant number 16-TF-KK].

Authors' contributions

BMA and GP conceived and designed the study. KK provided access to genomic data. BMA conducted bioinformatic assembly and analysis. BMA drafted the manuscript. All authors contributed to critically revising the manuscript. All authors read and approved the final version.

References

1. Koonin EV, Makarova KS, Aravind L. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol. 2001;55:709–42.

2. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008;9:605–18.

3. Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y, et al. Lateral transfers of large DNA fragments spread functional genes among grasses. Proc Natl Acad Sci U S A. 2019;116:4416–25.

4. Yang Z, Wafula EK, Kim G, Shahid S, McNeal JR, Ralph PE, et al. Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. Nat Plants. 2019. doi:10.1038/s41477-019-0458-0.

5. Stegemann S, Bock R. Exchange of genetic material between cells in plant tissue grafts. Science. 2009;324:649–51.

6. Kim G, LeBlanc ML, Wafula EK, dePamphilis CW, Westwood JH. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. Science. 2014;345:808–11.

7. Rice DW, Alverson 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;342:1468–73.

8. Richardson AO, Palmer JD. Horizontal gene transfer in plants. J Exp Bot. 2007;58:1–9.

9. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424:197–201.

10. Won H, Renner SS. Horizontal gene transfer from flowering plants to *Gnetum*. Proc Natl Acad Sci U S A. 2003;100:10824–9.

11. Koulintchenko M, Konstantinov Y, Dietrich A. Plant mitochondria actively import DNA via

585 the permeability transition pore complex. EMBO J. 2003;22:1245–54.

586 12. Logan DC. Plant mitochondrial dynamics. Biochim Biophys Acta. 2006;1763:430–41.

587 13. Arimura S-I, Yamamoto J, Aida GP, Nakazono M, Tsutsumi N. Frequent fusion and fission
588 of plant mitochondria with unequal nucleoid distribution. Proc Natl Acad Sci U S A.
589 2004;101:7805–8.

590 14. Davis CC, Xi Z. Horizontal gene transfer in parasitic plants. Curr Opin Plant Biol.
591 2015;26:14–9.

592 15. Petersen G, Anderson B, Braun H-P, Meyer EH, Møller IM. Mitochondria in parasitic
593 plants. Mitochondrion. 2020;52:173–82.

594 16. Wicke S, Naumann J. Molecular Evolution of Plastid Genomes in Parasitic Flowering
595 Plants. In: Chaw S-M, Jansen RK, editors. Advances in Botanical Research. Academic
596 Press; 2018. p. 315–47.

597 17. Krause K. From chloroplasts to “cryptic” plastids: evolution of plastid genomes in parasitic
598 plants. Curr Genet. 2008;54:111.

599 18. Petersen G, Cuenca A, Møller IM, Seberg O. Massive gene loss in mistletoe (*Viscum*,
600 Viscaceae) mitochondria. Sci Rep. 2015;5:17588.

601 19. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the
602 parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad*
603 genes. Proc Natl Acad Sci U S A. 2015;112:E3515–24.

604 20. Zervas A, Petersen G, Seberg O. Mitochondrial genome evolution in parasitic plants.
605 BMC Evol Biol. 2019;19:87.

606 21. Fan W, Zhu A, Kozaczek M, Shah N, Pabón-Mora N, González F, et al. Limited
607 mitogenomic degradation in response to a parasitic lifestyle in Orobanchaceae. Sci Rep.
608 2016;6:36285.

609 22. Sanchez-Puerta MV, Edera A, Gandini CL, Williams AV, Howell KA, Nevill PG, et al.
610 Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite
611 *Lophophytum mirabile* (Balanophoraceae). Mol Phylogenet Evol. 2019;132:243–50.

612 23. Sanchez-Puerta M, García LE, Wohlfeiler J, Ceriotti LF. Unparalleled replacement of
613 native mitochondrial genes by foreign homologs in a holoparasitic plant. New Phytol.
614 2017;214:376–87.

615 24. Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, et al. Massive mitochondrial
616 gene transfer in a parasitic flowering plant clade. PLoS Genet. 2013;9:e1003265.

617 25. Dawson JH, Musselman LJ, Wolswinkel P, Dörr I. Biology and control of *Cuscuta*.
618 Reviews of Weed Science. 1994;6:265–317.

619 26. van der Kooij TA, Krause K, Dörr I, Krupinska K. Molecular, functional and ultrastructural
620 characterisation of plastids from six species of the parasitic flowering plant genus *Cuscuta*.
621 Planta. 2000;210:701–7.

622 27. McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. Complete plastid genome sequences
623 suggest strong selection for retention of photosynthetic genes in the parasitic plant genus

Cuscuta. BMC Plant Biol. 2007;7:57.

28. 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.

29. Stefanović S, Olmstead RG. Down the slippery slope: plastid genome evolution in Convolvulaceae. J Mol Evol. 2005;61:292–305.

30. Revill MJW, Stanley S, Hibberd JM. Plastid genome structure and loss of photosynthetic ability in the parasitic genus Cuscuta. J Exp Bot. 2005;56:2477–86.

31. 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;9:2515.

32. Mower JP, Stefanović S, Young GJ, Palmer JD. Plant genetics: gene transfer from parasitic to host plants. Nature. 2004;432:165–6.

33. Mower JP, Stefanović S, Hao W, Gummow JS, Jain K, Ahmed D, et al. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol. 2010;8:150.

34. Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, et al. Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol. 2015;208:570–83.

35. Jiang L, Wijeratne AJ, Wijeratne S, Fraga M, Meulia T, Doohan D, et al. Profiling mRNAs of two *Cuscuta* species reveals possible candidate transcripts shared by parasitic plants. PLoS One. 2013;8:e81389.

36. Zhang Y, Fernandez-Aparicio M, Wafula EK, Das M, Jiao Y, Wickett NJ, et al. Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant *Phelipanche aegyptiaca* and related species. BMC Evol Biol. 2013;13:48.

37. Zhang D, Qi J, Yue J, Huang J, Sun T, Li S, et al. Root parasitic plant *Orobanche aegyptiaca* and shoot parasitic plant *Cuscuta australis* obtained Brassicaceae-specific strictosidine synthase-like genes by horizontal gene transfer. BMC Plant Biol. 2014;14:19.

38. Zhang Y, Wang D, Wang Y, Dong H, Yuan Y, Yang W, et al. Parasitic plant dodder (*Cuscuta* spp.): a new natural *Agrobacterium*-to-plant horizontal gene transfer species. Sci China Life Sci. 2020;63:312–6.

39. 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;9:2683.

40. Andrews S. FastQC. Babraham Bioinformatics. 2018. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed Feb 2019.

41. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory; 2014. https://sourceforge.net/projects/bbmap/. Accessed 2 Oct 2018.

42. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome

Res. 2017;27:722–36.

43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

44. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

45. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

46. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 2017;33:574–6.

47. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595.

48. 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;19:455–77.

49. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.

50. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016;32:2103–10.

51. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737–46.

52. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45:e18.

53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

54. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30:772–80.

55. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.

56. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 2010;27:1436–48.

57. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195.

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

59. Darzentas N. Circoletto: visualizing sequence similarity with Circos. Bioinformatics.

701 2010;26:2620–1.

702 60. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids
703 Res. 1999;27:573–80.

704 61. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1:
705 expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res.
706 2019;47:W59–64.

707 62. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2 – a
708 multiple sequence alignment editor and analysis workbench. Bioinformatics.
709 2009;25:1189–91.

710 63. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic
711 Acids Res. 2003;31:3406–15.

712 64. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC
713 Bioinformatics. 2005;6:96.

714 65. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved
715 genomic sequence with rearrangements. Genome Res. 2004;14:1394–403.

716 66. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation
717 of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
718 2011;7. https://www.embopress.org/doi/abs/10.1038/msb.2011.75.

719 67. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence
720 alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34 Web
721 Server issue:W609–12.

722 68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
723 large phylogenies. Bioinformatics. 2014;30:1312–3.

724 69. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and
725 evolutionary analyses in R. Bioinformatics. 2019;35:526–8.

726 70. Phanstiel DH, Boyle AP, Araya CL, Snyder MP. Sushi.R: flexible, quantitative and
727 integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics.
728 2014;30:2808–10.

729 71. Sawyer S. Statistical tests for detecting gene conversion. Mol Biol Evol. 1989;6:526–38.

730 72. Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes for
731 N-terminal targeting sequences. Proteomics. 2004;4:1581–90.

732 73. Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial
733 gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus
734 during angiosperm evolution. Proc Natl Acad Sci U S A. 2002;99:9905–12.

735 74. Kubo N, Arimura S-I. Discovery of the *rpl10* gene in diverse plant mitochondrial genomes
736 and its probable replacement by the nuclear gene for chloroplast RPL10 in two lineages of
737 angiosperms. DNA Res. 2010;17:1–9.

738 75. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to

739 the nucleus. Mol Phylogenet Evol. 2003;29:380–95.

740 76. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant
741 mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84:9054–8.

742 77. Laroche J, Li P, Maggia L, Bousquet J. Molecular evolution of angiosperm mitochondrial
743 introns and exons. Proc Natl Acad Sci U S A. 1997;94:5722–7.

744 78. Adams KL, Ong HC, Palmer JD. Mitochondrial gene transfer in pieces: fission of the
745 ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. Mol Biol
746 Evol. 2001;18:2289–97.

747 79. Liu S-L, Zhuang Y, Zhang P, Adams KL. Comparative analysis of structural diversity and
748 sequence evolution in plant mitochondrial genes transferred to the nucleus. Mol Biol Evol.
749 2009;26:875–91.

750 80. Petersen G, Cuenca A, Zervas A, Ross GT, Graham SW, Barrett CF, et al. Mitochondrial
751 genome evolution in Alismatales: size reduction and extensive loss of ribosomal protein
752 genes. PLoS One. 2017;12:e0177606.

753 81. Mower JP, Bonen L. Ribosomal protein L10 is encoded in the mitochondrial genome of
754 many land plants and green algae. BMC Evol Biol. 2009;9:265.

755 82. Gandini CL, Sanchez-Puerta MV. Foreign plastid sequences in plant mitochondria are
756 frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. Sci Rep.
757 2017;7:43402.

758 83. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The
759 alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet.
760 2019;15:e1008373.

761 84. Mower JP, Case AL, Floro ER, Willis JH. Evidence against equimolarity of large repeat
762 arrangements and a predominant master circle structure of the mitochondrial genome from a
763 monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol.
764 2012;4:670–86.

765 85. Mower JP, Sloan DB, Alverson AJ. Plant mitochondrial genome diversity: the genomics
766 revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. Plant Genome Diversity
767 Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics. Vienna:
768 Springer Vienna; 2012. p. 123–44.

769 86. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid
770 evolution of enormous, multichromosomal genomes in flowering plant mitochondria with
771 exceptionally high mutation rates. PLoS Biol. 2012;10:e1001241.

772 87. Marienfeld J, Unseld M, Brennicke A. The mitochondrial genome of Arabidopsis is
773 composed of both native and immigrant information. Trends Plant Sci. 1999;4:495–502.

774 88. de Longevialle AF, Small ID, Lurin C. Nuclearly encoded splicing factors implicated in
775 RNA splicing in higher plant organelles. Mol Plant. 2010;3:691–705.

776 89. Bonen L. *Cis*- and *trans*-splicing of group II introns in plant mitochondria. Mitochondrion.
777 2008;8:26–34.

778 90. Kim S, Yoon M-K. 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;56:177–88.

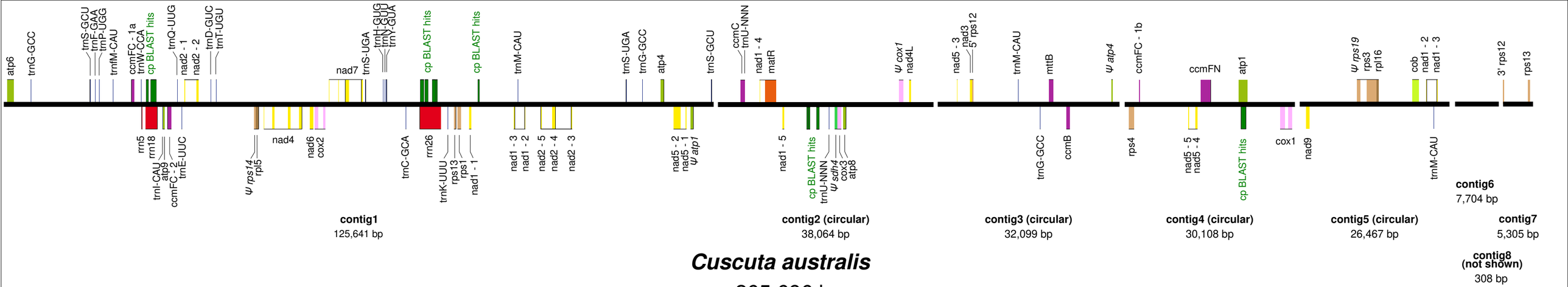
91. Glanz S, Kück U. *Trans*-splicing of organelle introns – a detour to continuous RNAs. Bioessays. 2009;31:921–34.

92. 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 $ccmF_C$ intron splicing and for the biogenesis of *c*-type cytochromes in *Arabidopsis* mitochondria: PORR domain mitochondrial splicing factor. Plant J. 2012;69:996–1005.

93. Giegé P, Rayapuram N, Meyer EH, Grienenberger JM, Bonnard G. $CcmF_C$ involved in cytochrome *c* maturation is present in a large sized complex in wheat mitochondria. FEBS Lett. 2004;563:165–9.

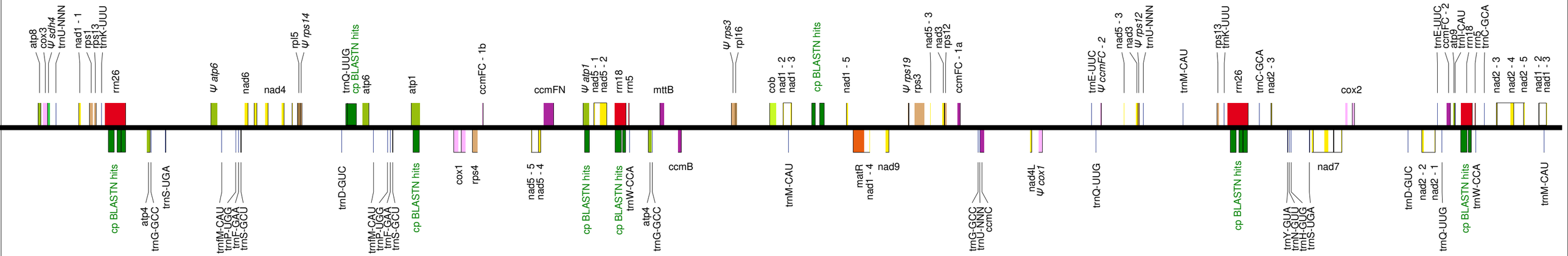
Figures

Fig. 1. Linear representations of mitogenomes of *Cuscuta australis* and *C. campestris*. The assembly of *C. australis* consists of eight contigs (divided by gaps), while that of *C. campestris* is represented as a single circle. Pseudogenes are indicated with a capital psi (Ψ).



Cuscuta australis

265,696 bp

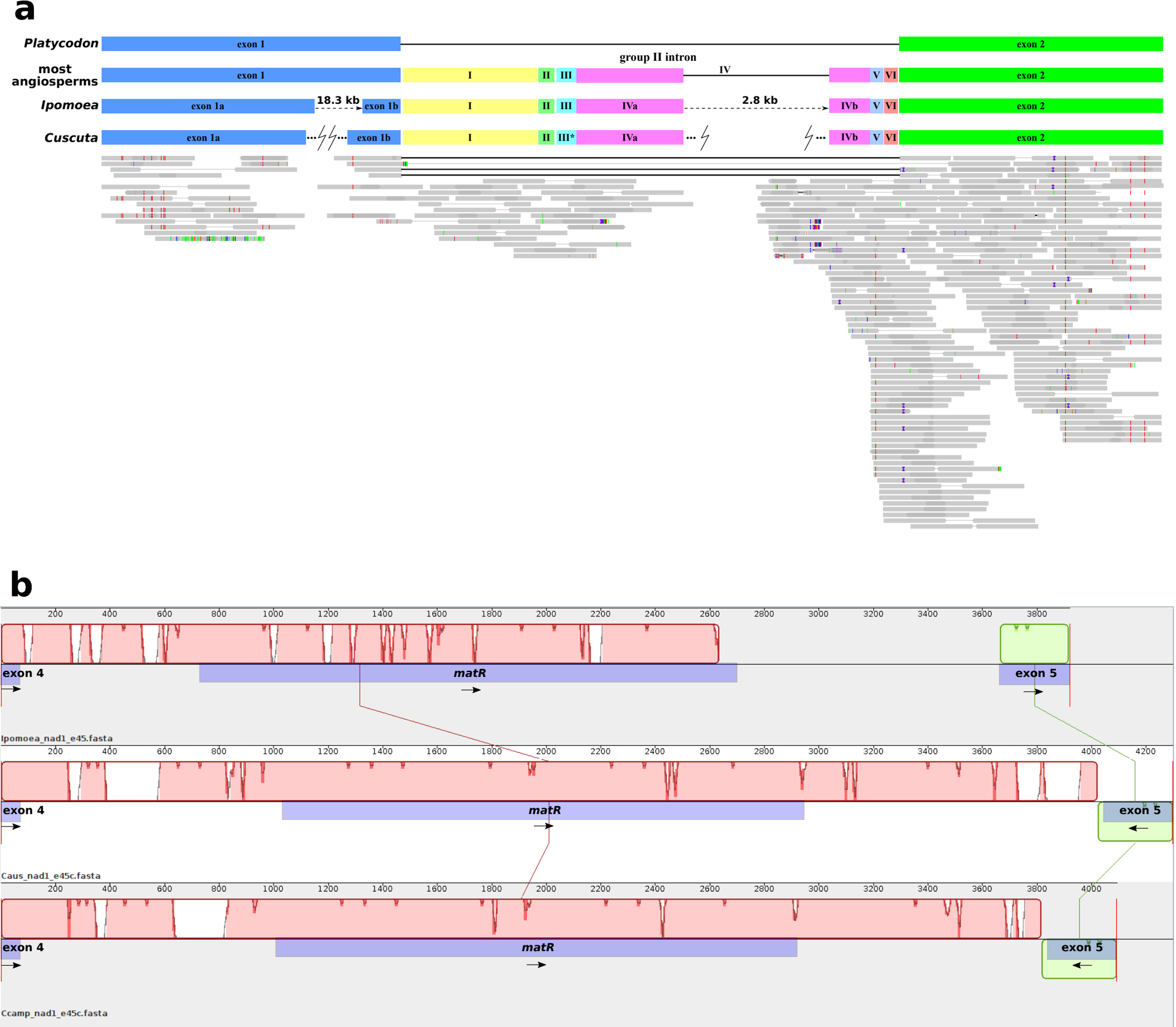


Cuscuta campestris

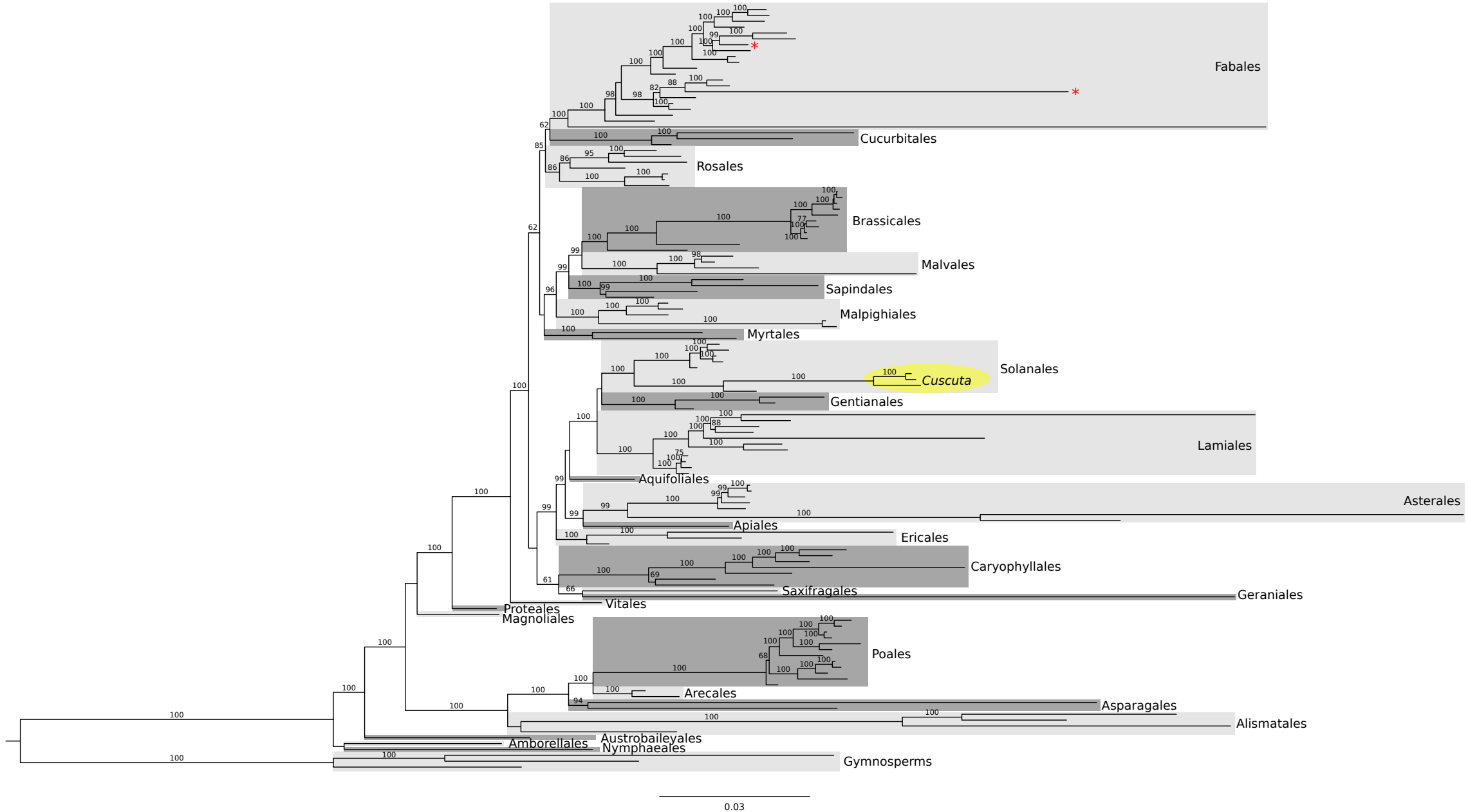
275,898 bp

10kb

797 Fig. 2. Notable structures for two mitochondrial genes in *Cuscuta*. (**a**) Schematic of the
798 *ccmFC* gene structure in angiosperms, and mapped mRNA reads to three DNA sequences
799 from *Cuscuta australis*. Group II intron domains are denoted with Roman numerals. Break
800 points are indicated with jagged symbols. Solid lines indicate connected sequences (no gap).
801 Dashed lines indicate intervening sequences. *Platycodon grandiflorus* does not have an
802 intron in its *ccmFC* gene. *The third domain in *Cuscuta* may not fold properly given its
803 sequence divergence. (**b**) Mauve alignment of the third portion of *nad1* (exon 4, *matR*, exon
804 5) for *Ipomoea nil*, *Cuscuta australis* and *Cuscuta campestris*. Coloured regions (red and
805 green) correspond to locally collinear blocks determined by Mauve. Shaded boxes with
806 labels show annotations, while arrows indicate the inferred direction of transcription.


a
Platycodon
exon 1
exon 2
group II intron
most angiosperms
exon 1
I
II
III
IV
V
VI
exon 2
Ipomoea
exon 1a
18.3 kb
exon 1b
I
II
III
IVa
2.8 kb
IVb
V
VI
exon 2
Cuscuta
exon 1a
exon 1b
I
II
III*
IVa
IVb
V
VI
exon 2
b
exon 4
matR
exon 5
Ipomoea_nad1_e45.fasta
exon 4
matR
exon 5
Caus_nad1_e45c.fasta
exon 4
matR
exon 5
Ccamp_nad1_e45c.fasta

807 Fig. 3. Maximum likelihood phylogenetic tree based on concatenation of all mitochondrial
808 genes present in *Cuscuta* together with representatives from angiosperm orders with fully
809 assembled mitogenomes. Orders are shaded, along with the outgroup gymnosperms, and
810 the location of *Cuscuta* is circled in yellow. Two asterisked branches in Fabales are (top to
811 bottom) a likely misidentified/contaminated *Bupleurum falcatum* (NC_035962.1) and
812 *Lophophytum mirabile*, with multiple horizontally transferred genes. Bootstrap support >60%
813 is shown above branches. The scale bar shows mean nucleotide substitutions per site.


Fabales
Cucurbitales
Rosales
Brassicales
Malvales
Sapindales
Malpighiales
Myrtales
Solanales
Cuscuta
Gentianales
Lamiales
Aquifoliales
Asterales
Apiales
Ericales
Caryophyllales
Saxifragales
Geraniales
Vitales
Proteales
Magnoliales
Poales
Arecales
Asparagales
Alismatales
Austrobaileyales
Amborellales
Nymphaeales
Gymnosperms
0.03