Mitochondrial genomes of two parasitic Cuscuta species lack clear evidence of horizontal gene transfer and retain unusually fragmented ccmF_C genes

Benjamin M. Anderson 1*, Kirsten Krause 2 and Gitte Petersen 1*

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 organelar 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 10 missing or pseudogenized genes often lost from angiosperm mitogenomes. Each mitogenome also possesses a structurally unusual ccmF_C 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 the Cuscuta mitogenomes, limiting the capacity to integrate foreign sequences.

Keywords: 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], though whether the inferred transfers are indeed foreign and/or functional has recently been questioned [25].

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 [26]. Photosynthetic ability is variably retained in the genus, and there is a diversity of chloroplast structure and plastome gene content [27–31]. For example, while some species are known to photosynthesize to some degree, carbon fixation in the bright yellow species C. campestris is below the compensation point [27] making it unable to survive and grow, let alone complete its life cycle, without its hosts. The range of possible host species for Cuscuta varies from narrow to especially broad with examples including members of Fabaceae, Apiaceae, Amaranthaceae, Asteraeae, Solanaceae, Linaceae, Ericaceae, Rutaceae and Amaryllidaceae among others [26]. The haustoria that Cuscuta plants use to obtain nutrients from their hosts are capable of both phloem and xylem connections [32] and are also capable of transmitting genetic material in the form of RNA [6] and DNA [4, 33], which together with the phylogenetically broad host range makes 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 [34, 35] and Geranium [36] 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 [37–39]. Nuclear genome assemblies coupled with transcriptomic work have also revealed numerous candidates for HGT in Cuscuta campestris [4, 33]. 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 [40]. 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 [41] and C. campestris [33] have been assembled, our knowledge of the mitogenome in Cuscuta is limited to a subset of genes from Cuscuta gronovii [36]. 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 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.

Results
Organelle genomes
The plastomes were assembled as single circular molecules having typical quadripartite structure with two inverted repeats separating small and large single copy regions (see Figs. S1 and S2, Additional file 1). The plastomes are 85,263 and 86,749 bp, with GC contents of 37.8 and 37.7% for C. australis and C. campestris, respectively. Beyond the inverted repeats (14,070 and 14,348 bp in C. australis and C. campestris, respectively), there are no dispersed repeats longer than 50 bp and with identity > 90%. Tandem repeats make up approximately 0.2 and 0.4% of the plastomes of C. australis and C. campestris, respectively. Similar to other reduced plastomes of Cuscuta species such as C. gronovii [29] and C. obtusiflora [28], 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 Table 1; Fig. S3, Additional file 1) 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. Read depth was largely consistent across the assemblies, except for drops at ends of contigs and in some locations possibly representing alternative conformations (see Fig. S4, Additional file 1). Recombinational activity as assessed with long repeats varied from 10 to 48% (average = 29%, n = 6) and 33–47% (average =
Table 1 Features of the mitogenomes of *Cuscuta australis*, *C. campestris* and their close relative *Ipomoea nil* (NC_031158)

| Feature                  | *C. australis* | *C. campestris* | *Ipomoea nil* |
|--------------------------|----------------|-----------------|---------------|
| Length (bp)              | 265,696        | 275,898         | 265,768       |
| GC content (%)           | 44.0           | 44.3            | 44.5          |
| Dispersed repeats (kb)   | 57.8           | 94.0            | 15.6          |
| Tandem repeats (kb)      | 13.2           | 6.89            | 0.67          |
| **Gene coding length / introns (bp)** | | | |
| *atp1*                   | 1530           | 1530            | 1530          |
| *atp4*                   | 579            | 579             | 579           |
| *atp6*                   | 1173           | 1173            | 1155          |
| *atp8*                   | 477            | 477             | 477           |
| *atp9*                   | 285            | 285             | 258           |
| *ccmB*                   | 627            | 627             | 627           |
| *ccmC*                   | 726            | 726             | 747           |
| *ccmF<sub>C</sub>₁<sub>a</sub> | 501            | 501             | 531 ("pseudo") |
| *ccmF<sub>C</sub>₁<sub>b</sub> | 128            | 128             | 92 ("pseudo") |
| *ccmF<sub>C</sub>₂<sub>a</sub> | 640            | 640             | 562 ("pseudo") |
| *ccmFN*                  | 1746           | 1755            | 1764          |
| *cob*                    | 1179           | 1179            | 1182          |
| *cox1*                   | 1584 / 509     | 1584 / 534      | 1584 / 971    |
| *cox2*                   | 1023 / 854     | 774 / 867       | 783 / 1343    |
| *cox3*                   | 798            | 798             | 798           |
| *matR*                   | 1911           | 1911            | 1968          |
| *mttB*                   | 774            | 774             | 840 ("pseudo") |
| *nad1<sub>1</sub>*      | 386            | 386             | 385           |
| *nad1<sub>2</sub>–<sub>3</sub>* | 274 / 1772     | 274 / 1223      | 275 / 1445    |
| *nad1<sub>4</sub>–<sub>5</sub>* | 318 / 3982     | 318 / 3776      | 318 / 3604    |
| *nad2<sub>1</sub>–<sub>2</sub>* | 546 / 1941     | 546 / 1870      | 545 / 1233    |
| *nad2<sub>3</sub>–<sub>5</sub>* | 921 / 2785, 1833 | 921 / 2348, 1773 | 922 / 2417, 1455 |
| *nad3*                   | 357            | 357             | 357           |
| *nad4*                   | 1488 / 1478, 2359, 1398 | 1488 / 1498, 2390, 1426 | 1488 / 1418, 3200, 2827 |
| *nad4L*                  | 303            | 303             | 300           |
| *nad5<sub>1</sub>–<sub>2</sub>* | 1446 / 838     | 1446 / 857      | 1446 / 844    |
| *nad5<sub>3</sub>*       | 22             | 22              | 22            |
| *nad5<sub>4</sub>–<sub>5</sub>* | 542 / 1088     | 542 / 1175      | 542 / 1339    |
| *nad6*                   | 618            | 618             | 618           |
| *nad7*                   | 1176 / 1553, 1124, 2144 | 1176 / 1294, 910, 1957 | 1185 / 912, 1468, 1951 |
| *nad9*                   | 573            | 573             | 573           |
| *rpl2*                   | –              | –               | –             |
| *rpl5*                   | 555            | 555             | 555           |
| *rpl10*                  | –              | –               | 489           |
| *rpl16*                  | 516            | 456             | 516           |
| *rps1*                   | 612            | 612             | 597           |
| *rps2*                   | –              | –               | –             |
| *rps3*                   | 1830 / 1057    | 1794 / 996      | 1671 / 1510   |
| *rps4*                   | 912            | 909             | 933           |
37%, n = 12) for *C. australis* and *C. campestris*, respectively (see Tables S1 and S2, Additional file 1).

For each mitogenome, we annotated 31 protein-coding genes, 3 ribosomal RNA 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 trnT-UGU is missing. 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*.

The *ccmF* gene was problematic to annotate in both species (annotated as a pseudogene in *C. gronovii* and *Ipomoea nil*) and may not be properly transcribed, as it appears to be both trans-spliced and missing a portion of the first exon. Based on alignments with other angiosperms, including the intron-less *Platycodon grandiflorus* (NC_035958), it appears that the *ccmF* gene is split across three locations in both of the *Cuscuta* mitogenomes (Fig. 2). A large portion of exon 1 is separated

| Feature | *C. australis* | *C. campestris* | *Ipomoea nil* |
|---------|---------------|----------------|--------------|
| rps7    | -             | -              | -            |
| rps10   | -             | -              | 363 / 902    |
| rps11   | -             | -              | -            |
| rps12   | 378           | 378            | 378          |
| rps13   | 351           | 351            | 351          |
| rps14   | 276 (pseudo)  | 258 (pseudo)   | 303          |
| rps19   | 209 (pseudo)  | 114 (pseudo)   | 285          |
| sdh3    | -             | -              | -            |
| sdh4    | 431 (pseudo)  | 486 (pseudo)   | 408          |
| rrs5    | 116           | 116            | 119          |
| rrs18   | 2075          | 2037           | 1943         |
| rrs26   | 3753          | 3656           | 3377         |
| trnC-GCA| 71            | 71             | 71           |
| trnD-GUC| 74            | 74             | 74           |
| trnE-UUC| 72            | 72             | 72           |
| trnF-GAA| 74            | 74             | 74           |
| trnG-GCC| 72            | 72             | 72           |
| trnH-GUG| 74            | 74             | 74           |
| trnI-CAU| 74            | 74             | 74           |
| trnK-UUU| 73            | 73             | 73           |
| trnM-CAU| 73            | 73             | 73           |
| trnM-CAU| 73            | 73             | 73           |
| trnN-GUU| 72            | 72             | 72           |
| trnP-UGG| 75            | 75             | 74           |
| trnQ-UGG| 72            | 72             | 72           |
| trnQ-UUG| -             | 72             | -            |
| trnS-GCA| 88            | 88             | 88           |
| trnS-GGA| -             | -              | 88           |
| trnS-UGA| 87            | 87             | 87           |
| trnT-UGU| 72            | 72             | -            |
| trnW-CCA| 74            | 74             | 74           |
| trnY-GUA| 83            | 83             | 83           |

Table showing a summary of the three mitogenomes, including gene content. Protein-coding, ribosomal RNA and transfer RNA genes often found in angiosperm mitogenomes are listed along with their lengths, the lengths of their introns, and whether they are missing (‘-’) or pseudogenized (‘pseudo’).
from the 3′ end of exon 1 and most of the group II intron. The remainder of the intron plus exon 2 are in a third location. Mapping mRNA reads resulted in low coverage of the ccmFC gene, but provided some support for successful trans-splicing of the 3′ end of exon 1 and exon 2 (see Fig. 2). There was no indication of an mRNA product that included both portions of exon 1, however, making it unclear whether the majority of exon 1 is actually translated into part of a functional product. PREP-MT predicted 14 editing sites, three of which didn’t have transcriptome support, and missed four sites that appeared to be RNA edited based on transcriptome read mapping (see Fig. S5, Additional file 1). Mapping mRNA reads against other mitochondrial gene sequences (e.g. atp6, ccmC; data not shown) recovered similarly low read depths as the mapping to ccmFC, suggesting read depth in this case is unreliable for effectively determining relative expression of mitochondrial genes. For accurate investigation of mitochondrial transcription of the genes of interest, other techniques like mitochondrial run-on transcription assays [42] would be more suitable.

Another odd gene structure found in both Cuscuta species is the final portion of nad1, namely exon 4, matR and exon 5 (see Fig. S6 Additional file 2). In the two Cuscuta assemblies, exon 5 is found in the reverse orientation on the opposite strand, while exon 4 is located on the same strand as matR. This may indicate nad1i728 has shifted to trans-splicing (as seen in some other angiosperms, e.g. Nicotiana tabacum, NC_006581.1), but if that is the case, it seems odd that exon 5 is in such close proximity to the rest of the intron and exon 4.

**Potential intracellular transfers**

Using reference plastomes, we plotted plastid hits on the mitogenome assemblies (see Figs. S7 and S8, Additional file 2). Excluding hits that occurred in the ribosomal genes or atp1, which are similar to plastid ribosomal genes and atpA, we found 8 regions with hits, representing approximately 0.6 and 1.1% of the mitogenomes of C. australis and C. campestris, respectively (Table 2).

Phylogenetic trees for hits 2 and 7 (see Fig. S9, Additional file 2) support the mitochondrial sequences grouping with Cuscuta or Ipomoea chloroplast sequences, consistent with intracellular transfer. Both species share two regions corresponding to HMMER hits to plastid genes ndhB and psbB (hits 4 and 5). Phylogenetic analysis of hits 4 and 5 are partly unresolved but suggestive of potential HGT, as the two Cuscuta species lack intact copies of ndhB in their plastomes (hit 4) and...
the psbB hits group closer to Malpighiales than other Cuscuta chloroplast sequences. In both Cuscuta species, the psbB and ndhB hit regions are within 2 kb of each other, possibly indicating a larger transferred region. Bootstrap support is low in the ndhB tree, with taxa from asterids and rosids, making it unclear whether the two hits are associated with the same potential donor lineage.

Searches of the nuclear genomes produced markedly different results for the two Cuscuta species. In C. australis, the ~260 Mbp assembly hit the mitochondrial assembly for cumulatively ~27% of its length, but in C. campestris, the ~477 Mbp assembly hit the mitochondrial assembly for ~99.8% of its length or essentially the entire assembly. Three scaffolds with lengths of 172, 43 and 21 kb are essentially entirely mitochondrial. These scaffolds could tentatively indicate recent large-scale integration of mitochondrial sequences into the nuclear genome, but could also be simple assembly artifacts since mitochondrial contaminations are difficult to identify.

| Mitochondrial gene | Nuclear copy length (bp) | Reference average length (bp) | Putative origin |
|--------------------|--------------------------|-------------------------------|----------------|
| rps2               | 618                      | 978                           | vertical; nuclear copy or UTP--glucose-1-phosphate uridylyltransferase |
| rps10              | 387                      | 354                           | unclear; possible HGT |
| rps11              | 432                      | 486                           | vertical; nuclear copy |
| rps14              | 321                      | 295                           | vertical; nuclear copy |
| rps19              | 273                      | 286                           | unclear; possible HGT |
| sdh3               | 321, 324                 | 343                           | vertical; nuclear copy |
| sdh4               | 288                      | 418                           | unclear; possible HGT |

Table 3 Nuclear candidate copies of missing mitochondrial genes in Cuscuta australis and C. campestris

Table showing the nuclear candidate copies, including the lengths of reference mitochondrial genes. The putative origin is based on phylogenetic analysis of top NCBI BLASTN hits together with reference mitogenomes
Fig. 3 (See legend on next page.)
Potential candidates for missing mitochondrial genes were identified in both nuclear genomes for rps2, rps10, rps11, rps14, rps19, sdh3 and sdh4 (Table 3). All the candidates had intact open reading frames except for sdh3, which had no determined start codon. Phylogenetic analysis of the candidates with copies from the reference mitogenomes and top NCBI BLASTN hits consistently put the nuclear copies on longer branches relative to the mitogenome copies (see Figs. S10–13, Additional file 2). In some cases, the trees suggest vertical inheritance of a nuclear copy (rps2, rps11, rps14, sdh3) but in three cases it is unclear and could represent HGT (rps10, rps19, sdh4), though long branch attraction may obscure phylogenetic relationships. Searching for indications of mitochondrial targeting sequences with Predotar v. 1.04 [43] indicated possible mitochondrial targeting (p 0.21–0.44) for rps10, rps14, rps19 and sdh3 (if an ‘M’ is added in place of the undetermined start codon).

Horizontal gene transfer

Phylogenetic analysis of the concatenation of all mitochondrial genes present in Cuscuta together with representatives from angiosperm orders with fully assembled mitogenomes. Orders are indicated at the right along with the outgroup gymnosperms, and the location of Cuscuta is highlighted in yellow. Bootstrap support of 60% or greater is shown above branches. The sequence for Bupleurum falcatum (NC_035962.1) is likely misidentified.

| Fig. 3 | Maximum likelihood cladogram based on concatenation of all known mitochondrial genes present in Cuscuta together with representatives from angiosperm orders with fully assembled mitogenomes. Orders are indicated at the right along with the outgroup gymnosperms, and the location of Cuscuta is highlighted in yellow. Bootstrap support of 60% or greater is shown above branches. The sequence for Bupleurum falcatum (NC_035962.1) is likely misidentified. |

Recentley sequenced genomes of parasitic Cuscuta species revealed extensive HGT events into their nuclear genomes from a broad range of non-parasitic angiosperms that serve as their hosts. It was a reasonable expectation that there would be high transfer activity of genetic material into the mitogenomes of these parasites as well. The organellar genomes of the two Cuscuta species that we have assembled and annotated here, however, are not clearly different in gene content from other Cuscuta plastomes or typical angiosperm mitogenomes. Based on phylogenetic analysis, we found little evidence for HGT into the mitogenomes of the two species, an unexpected result given the parasitic lifestyle of Cuscuta and evidence for HGT candidates in their nuclear genomes. In the process of examining the mitogenomes, we also identified a novel structure of the ccmF<sub>C</sub> gene in Ipomoea and Cuscuta, indicative of a shift to transsplicing of the group II intron and possibly a shorter product.

Gene losses and possible intracellular transfers

All of the 10 genes missing from the two Cuscuta mitogenomes are commonly found to be lost during angiosperm evolution [44, 45]. That rps2 and rps11 are missing is consistent with the inference of their ancient losses in angiosperms [44]. For 7 of the 10 missing or pseudogenized genes, potential candidates for nuclear copies were found. Transfers of mitochondrial genes to the nucleus have been inferred to be numerous across plant evolution and seem to be the primary mode of functional gene loss from the mitogenome [46]. Whether the gene products from the nuclear candidates are functional or imported into the mitochondrion is uncertain, but tests for targeting sequences suggested some of them might be targeted to the mitochondrion. The nuclear candidates are on longer branches in phylogenetic trees.
compared to the mitochondrial copies, suggesting higher substitution rates. This is consistent with typical observations of higher substitution rates in the nuclear genome compared to the mitogenome [47], and of higher rates in nuclear copies of genes transferred from the mitogenome compared to the mitogenome copies [48], though exceptions exist, such as in Silene [49].

The three missing mitochondrial genes not detected in the nuclear genomes are rpl2, rpl10 and rps7. Losses of rpl2 and rps7 are frequent among angiosperms, but only for rpl2 are losses commonly associated with functional transfers to the nucleus [36, 44, 50, 51]. Losses of rpl10 are less common and have been associated with functional replacement by a duplicated nuclear gene normally encoding chloroplast RPL10 in some monocot and Brassicaceae lineages [45, 52, 53]. Searching ORFs from both nuclear genomes for possible copies of RPL10 suggested the presence of one (C. australis) or two (C. campestris) genes more similar to the chloroplast-targeted copy from Arabidopsis (NP_196855) than to the mitochondrial-targeted one (NP_187843). Targeting predictions from Predotar for the copies in Cuscuta provided no support for targeting the chloroplast or the mitochondrion.

Plastid transfers appear to be limited in the mitogenomes, with a 148 bp hit in C. australis and a 1.6 kb hit in C. campestris. The other main candidate regions included hits to psbB and ndhB, but based on phylogenetic analysis appear to be more similar to regions in distantly related plants. The trees are inconclusive as to the origin of the region, but suggest a horizontal transfer, possibly via an intracellular plastome transfer to the donor’s mitogenome and then a plant to plant transfer between mitogenomes, a common route for plastome to mitogenome HGT [54]. The close association of the psbB and ndhB hits might indicate a transfer of a larger contiguous block of DNA from a donor with the genes in close proximity, but this is speculative and the donor lineage is not clear from our search and trees.

**Lack of horizontal gene transfer**

Phylogenetic evidence for HGT in the 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. BLASTN-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. It is possible they represent vertically-inherited mitogenomic sequences. Given the dynamic nature of mitogenome size and structure in angiosperms [49, 55–57], 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 the reference set of mitogenomes (with 168 angiosperms), the average size of the angiosperm mitogenomes is roughly 670 kb, while the assemblies are less than half that. Given there has been a correspondence observed between mitogenome size and the amount of incorporated plastid transfers [54] and that plant mitochondrial genomes appear to have relaxed restriction on the size of intergenic spacers (e.g. [58]), 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. Other explanations could include population sizes and demographic effects that make it more or less likely to retain incorporations of foreign DNA in a species.

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 [49]. In a study of the massive differences in size amongst Silene mitogenomes, Sloan et al. [49] found that the larger mitogenomes experienced lower recombinational activity (< 20%) and had more divergent repeats, suggesting reduced gene conversion. The 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 ccmF_C and its trans-splicing intron**

Both Cuscuta species and their close autotrophic relative Ipomoea nil appear to share a different structure of their
ccmF_C 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 proper splicing [59]. While the exact mechanism of trans-splicing for group II introns and the splicing factors required remain unclear, this type of splicing has been observed in a number of other angiosperm mitochondrial genes, namely nad1, nad2 and nad5 [60] and cox2 [61]. Typically, the break point for the two pieces of the intron that are trans-spliced is in domain IV [60, 62], which is where the intron is broken in Cuscuta and Ipomoea. The evidence presented here, namely the mRNA reads that indicate a product joined from two disparate locations in the genome and the break in domain IV, suggests this is the fifth such example of a gene with a group II intron that has shifted to trans-splicing.

Angiosperm ccmF_C genes examined so far typically have intact rather than broken group II introns. For splicing of the intact intron, an essential nuclear-encoded splicing factor, WTF9, has been identified which shares an RNA-binding domain with another factor involved in intron splicing in plastids [63]. In examining the role of WTF9, des Francs-Small et al. [63] also elucidated the role of ccmF_C in cytochrome biogenesis, showing that proper splicing of the intron for the formation of a complete protein was essential to the generation of cytochromes c and c_1 in Arabidopsis. In addition, splicing-impaired mutants had higher levels of alternative oxidase, suggesting problems with the respiratory chain [63]. It appears that being able to produce a fully intact ccmF_C product is important for respiratory function, which makes the break in the first exon in Cuscuta and Ipomoea puzzling. While there is evidence that the intron is trans-spliced and connects the 3’ portion of exon 1 with exon 2, there is no indication that the larger 5’ portion of exon 1 is connected for translation into the final product. Transcriptomic reads show RNA editing in all parts of the gene (see Fig. 2 and Fig. S4, Additional file 1), similar to predictions from PREP-MT, but whether this indicates an intact product is unclear. Pseudogenization of ccmF_C has only been reported for Silene conica amongst angiosperms [49], though there was indication that a truncated form of the protein might still be expressed and functional. Normally, the full ccmF_C product is thought to be integrated into a larger complex potentially involved in heme ligation [64]. Another possible explanation for the process might be a continued trend of gene breakup, similar to the fission of ccmF_C in Trifolium, Brassicaceae and Allium [65]. The functionality of at least part of the ccmF_C gene would be consistent with its conservation in both Ipomoea and Cuscuta, though their copies are divergent relative to other members of Solanales (see Fig. S14, Additional file 3). Gene loci coding for homologs of the WTF9 splicing factor are present in C. campestris and C. australis (data not shown), excluding this as a reason for the switch. How Cuscuta and Ipomoea are able to function effectively with their ccmF_C structure is unclear, and additional work would be helpful to characterise and evaluate the ccmF_C protein products in this group of plants.

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

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

Methods
We downloaded from GenBank portions of previously published datasets from whole genome sequencing studies of Cuscuta australis and C. campestris [33, 41]. 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 [66] 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. 3.80 [67; https://sourceforge.net/projects/bbmap/]. As part of initial assembly, long reads were corrected then trimmed using default settings in Canu v. 1.8 [68]. 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* [41], 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). Transcriptomic reads for *Ipomoea nil* were obtained from three sources: the 1KP project [69], ERR2040616; a transcriptome study [70], SRR1772255; and a genome project [71], DRR024544–9. Reads were also deduplicated and trimmed for adapters and low quality bases, with 157 M reads retained with an average read length of 97 bp (81% 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 BLASTN v. 2.6.0+ [72–74] with default parameters for hits to reference genomes (see Tables S3 and S4, Additional file 4), 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 [75] 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 500–25,000 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 [76], which when given only short reads optimizes a SPAdes v. 3.14.0 [77] assembly.

Long read datasets were assembled in Canu v. 1.8 [68] 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 [78] to identify those with BLAST (blastn megablast default parameters) hits to the reference genomes and which had similar depths (for distinguishing plastid and mitochondrial contigs). We considered as candidates contigs with >50% coverage of hits to the respective references, and in cases of less than 50% coverage, whether read depth was similar to depth of contigs having greater coverage of hits. 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 [79] and Racon v. 1.4.3 [80] 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 the assembly approach, we also used NOVOPlasty v. 3.7.2 [81] 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 locations, but this differed between assemblies. For *C. australis*, a second round of Unicycler was run after extending contigs by up to 100 kbp using Tadpole (part of the BBMap package) with settings el = 100,000 er = 100,000 mode = extend k = 31 and the full short read dataset, and then mapping short reads again. Simplifying the assembly graph included removing small contigs that were fully present in a larger contig or were too short (e.g. < 100 bp). We then checked long read contigs from the initial assembly in Bandage to ensure they were covered by hits from the remaining assembly. Since there were four linear and four circular contigs, and since the circular contigs did not share repeats that could be used to merge them, we were unable to merge all contigs into a single contig. Repeats are present (though there are no long repeats shared between circles) and there are alternative paths through the assembly, so our representation is one of multiple possible and does not reflect the actual structure of the mitochondrial genome in the plants. While we were not able to merge the contigs in *C. australis*, we could join circular contigs for the assembly of *C. campestris*. For *C. campestris*, the first round of Unicycler produced four circular contigs along with three small linear contigs (linear totalling < 5 kb) which hit the
plastome or had low depth and were absent from the long read contigs. The short linear contigs were not included in further assembly. The four circles had shared long repeats, and these were examined using long reads to determine if multiple configurations were present, suggesting the separate contigs are sometimes contiguous. Long reads were mapped and aligned to extracted repeats with Minimap2 v. 2.14-r883 [79], Samtools v. 1.7 [82] and MAFFT v. 7.310 [83] and using a custom Python script including ‘pysam’ (https://github.com/pysam-developers/pysam). When the reads indicated multiple configurations across the repeat, we manually joined the circles. The circles all shared repeats that showed multiple configurations, so we were able to merge the four circles into a single circular contig.

**Read mapping and annotation**

In order to check the mitogenome assemblies for clear errors, we conservatively mapped short and long reads then manually compared the resulting BAM files in IGV v. 2.6+ [84]. Short reads were mapped using BBMap with settings pairedonly = t minid = 0.9 ambiguous = random maxindel = 100 strictmaxindel = t pairlen = 1000. Long reads were mapped using Minimap2 in mode ‘map-pb’ and filtered with Samtools with option -F 2308, followed by filtering the resulting BAM file with a custom Python script using ‘pysam’ to only retain mappings where reads covered the target for at least 80% of their length. Long reads were used to check structural assembly and cases of insertions, where high coverage plastid reads can lead to misassembly of the mitogenomic sequence. The long reads passing through an insertion can be used to identify the lower-coverage short reads that are mitochondrial and allow for manual correction. For *C. australis*, only one location suggested a small (28 bp) sequence that needed to be deleted (absent in most long reads). For *C. campestris*, one location of a plastome insert needed three single base pair corrections and a small (9 bp) deletion to match the long reads that spanned the region rather than the higher depth plastid reads.

Plastomes and mitogenomes were annotated using custom scripts to BLAST (BLASTN, BLASTX, default search settings) against nucleotide and protein databases built using reference plastid (14 from 13 genera and 13 families, see Table S3, Additional file 4) and mitochondrial (172 from 130 genera and 62 families, see Table S4, Additional file 4) genomes, an approach similar to that used in Mitofy [85]. For mitogenomes, we also used the references to build protein alignments for constructing a HMMER v. 3.3 (http://hmmer.org) database that we used to search open reading frames (ORFs) found using ORFfinder v. 0.4.3 (NCBI; https://www.ncbi.nlm.nih.gov/orffinder/). Following searches, hits were examined and adjusted manually, with input from MFAnnot v. 1.35 (https://github.com/BFL-lab/MFannot) to select intron boundaries. Transfer RNAs were also detected and checked with tRNAscan-SE 2.0 [87]. Pseudogenes were subjectively annotated if hits had clear frameshifts and/or internal stop codons, and gene copies broken by repeats or alternative paths in the assembly were annotated as fragments.

Dispersed repeats were detected by searching the organelar genomes against themselves using BLASTN with an E value of 1e-10 and filtering out hits less than 50 bp and lower than 90% identity. Overlapping hits were then filtered with a custom python script, and the percentage repeat content calculated. For visualization, the genomes were searched against themselves using Circos [88] with an E value chosen to approximate a 90% identity threshold (1e-70 and 1e-40 for *C. australis* and *C. campestris*, respectively). As a proxy for recombinational activity, we used long repeats (> 500 bp, > 90% identity) and mapped long reads. A random selection of up to 50 long reads that mapped to each repeat were aligned with the repeat and evaluated manually for whether they showed indications of incongruence on either side of the repeat. The proportion of the mapped reads showing incongruence was used as a proxy for recombinational activity. Tandem repeat content was determined using Tandem Repeat Finder v. 4.09 [89] with recommended settings: 2 7 7 80 10 50,500. Organelle genome figures were based on output from OGDRAW [90].

To investigate the transcription and splicing of the mitochondrial *ccmF* 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 *ccmF* locations in the *C. australis* assembly. We also aligned *ccmF* DNA sequences from both *Cuscuta* species with *ccmF* 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 [91] and IGV v. 2.8.0 [84]. Intron domains were roughly determined by iteratively running the Mfold web server [92] using the ccmFci829 from *Nicotiana tabacum* (NC_006581). RNA editing sites were predicted with PREP-MT [93] to compare with those observed in the transcriptomic reads. We also aligned *Ipomoea nil* transcriptomic sequences to similar *ccmF* DNA sequences from its mitogenome to compare with *Cuscuta*.

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 [94]. We also mapped transcriptomic reads from *C. australis* to *nad1* exons separated by 100 Ns to evaluate splicing and whether there was evidence for an intact product.
Potential intracellular transfers
To detect the presence of transfers from a plastome to the mitogenome, we searched each mitogenome assembly against the reference plastomes (Table S3, Additional file 4) using BLASTN with default parameters. We also searched the mitochondrial ORFs against a HMMER database built with the reference plastomes to detect whether there were intact portions of plastid genes present and to interpret blast hits. Broader searches used the NCBI non-redundant nucleotide database (https://blast.ncbi.nlm.nih.gov/). We evaluated plastome hits of interest phylogenetically using top NCBI results in RAxML v. 8.2.12 [95] with rapid bootstrapping and a search for best tree in a single run, opting for 1000 bootstrap replicates under the GTRGAMMA model.

To determine whether any missing mitochondrial genes were present in the nuclear genome, and whether there were other transfers between the mitogenome and the nuclear genome, we conducted BLAST searches of the assemblies and of the reference mitogenomes (Table S4, Additional file 4) against the two published nuclear genomes (C. australis: GCA_003260385.1, C. campestris: PRJEB19879). For assessment of total hit coverage of the assemblies, we ignored hits <250 bp and with <80% identity. For searches for missing genes, we ignored hits with <50% identity.

Horizontal gene transfer
We investigated HGT using a phylogenetic approach, in which we built phylogenetic trees for each mitochondrial gene present in Cuscuta (as well as one based on all genes concatenated) and examined whether Cuscuta sequences were supported as grouping with their close relatives (particularly Ipomoea nil) or with more distantly related plants. Using reference mitogenomes (see Table S4, Additional file 4) and the Cuscuta assemblies along with data from C. gronovii (GenBank accessions KP940494–514), we extracted all CDS protein translations and corresponding nucleotide sequences, as well as nucleotide sequences for ribosomal RNAs. We removed some sequences in cases of high divergence (e.g. for Viscum scurruloidesum) and when alignments suggested problematic annotations, and in some cases we had to reverse complement ribosomal RNA sequences. For each gene, we created protein alignments with ClustalOmega v. 1.2.4 [96] using default settings. We then converted the protein alignments back to nucleotide alignments with PAL2NAL v. 14 [97] using default settings. Alignments were run through PREP-MT [93] to predict editing sites, with sites having a prediction score of > = 0.5 removed prior to analysis. Alignments were checked manually for apparent misalignments and problems with annotation around intron boundaries, and the cox1 co-conversion tract [98] was removed. For ribosomal RNA sequences, we aligned with MAFFT v. 7.310 using options --localpair and --maxiterate 1000. For each alignment we used a custom python script to eliminate positions with >50% missing data, then ran RAxML v. 8.2.12 [95] with rapid bootstrapping and a search for best tree in a single run, opting for 1000 bootstrap replicates under the GTRGAMMA model. Finally, we plotted the resulting trees using R package ape [99], rooting them on non-angiosperm seed plants, and manually examined whether Cuscuta sequences grouped together, with their close relative Ipomoea nil, and with Solanales or whether they grouped elsewhere with bootstrap support >80%

To explore intergenic sequences, we searched the reference mitogenomes against the Cuscuta assemblies using BLASTN with default parameters, removing hits <250 bp and with <80% identity. For clearer display, we used custom scripts to filter BLAST hits so that hits of equivalent length (within 50 bp) or less and more than 2% lower identity than top hits were removed along the assembly. We plotted hits using the R package Sushi v. 1.26.0 [100]. We then manually examined hits in intergenic regions, looking for where the top hits were not from Solanales and evaluating whether the hits were suggestive of longer, potentially transferred tracts.

In cases of possibly chimeric gene sequences based on top blast hits, we assessed evidence for conversion using GENECONV v. 1.81 [101].

Abbreviations
CDS: Coding sequence; HGT: Horizontal gene transfer; ORF: Open reading frame.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-08105-z.

Additional file 1: Fig. S1. Chloroplast genome of Cuscuta australis. Transcription is counterclockwise around the outside of the circle and clockwise inside. The grey inner graph is GC content. Abbreviations: LSC = large single copy region, SSC = small single copy region, IRA/B = inverted repeat A/B. Fig. S2. Chloroplast genome of Cuscuta campestris. Transcription is counterclockwise around the outside of the circle and clockwise inside. The grey inner graph is GC content. Abbreviations: LSC = large single copy region, SSC = small single copy region, IRA/B = inverted repeat A/B. Fig. S3. Dispersed repeats in the mitogenomes of Cuscuta australis (left) and C. campestris (right). Generated with Circoletto (http://tools.bat.infspire.org/circoletto/). Fig. S4. Read depth for mitogenome assemblies of Cuscuta australis (top) and C. campestris (bottom). Short read depth is in black while long read depth is in green. Breakpoints between contigs in C. australis are blue lines. Minimum and mean read depths are indicated, with colours of text matching respective read depth. Fig. S5. Edited sites in comFC from Cuscuta australis based on mapped transcriptomic reads and predicted by PREP-MT. Exons are indicated above the mapped reads, with circles corresponding to panels above showing the mismatches in more detail. Positions along the alignment are numbered, with positions not predicted by PREP-MT circled in red. The table shows details of the results from PREP-MT. Table S1. Repeats recombination analysis for Cuscuta australis. The analysis is based on non-
overlapping repeats > 500 bp and > 90% identity (blastn eval 1e-10), and mapping long reads to those repeats. Table S2. Repeats recombination analysis for Cuscuta campestris. The analysis is based on non-overlapping repeats > 500 bp and > 90% identity (blastn eval 1e-10), and mapping long reads to those repeats.

Additional file 2: Fig. S6. Schematic of the structure of nad1 in Cuscuta australis based on DNA alignment with C. campestris and Ipomoea nil and alignment of transcriptomic reads. At top, a Mauve alignment of the section from exon1 to exon5 (containing nad1) is shown for the three species (from top to bottom: Ipomoea nil, Cuscuta australis, C. campestris). Direction of transcription is shown with arrows. Exon structure is shown below with mRNA reads aligned to the DNA sequences separated by 100 Ns. Black lines demarcate the exonic sequences. Coloured lines in aligned reads indicate mismatches, corresponding to putative editing sites in most cases. Figs. S7 and S8. BLASTN hits for reference mitogenomes and plastomes plotted along the mitogenomes of Cuscuta australis and C. campestris. Plastome hits are shown in green along the linear representations of the mitogenomes, while hits to reference plastomes are shown by order above. The percentage identity of the mitogenomic hits is coloured based on the legend at the right. Solanales is shaded, as the order to which Cuscuta belongs. Hits > 500 bp in regions of known plastid genes are noted with Solanales circled. Possible organelle hits are boxed. Fig. S9. Phylogenetic trees of plastid hits in mitogenomes of Cuscuta australis and C. campestris, analysed with top BLASTN hits from NCBI. All support values from 1000 rapid bootstrap replicates are shown above branches. Hits correspond to Table 2 and include the location within each mitogenome. Cuscuta sequences from plastomes are highlighted in green, while Cuscuta mitogenomic sequences are highlighted in yellow. Scale bars indicate inferred nucleotide substitutions per site. Figs. S10, S11, S12 and S13. Phylogenetic trees for nuclear candidate copies of mitochondrial genes in Cuscuta australis and C. campestris in relation to known mitochondrial copies in other plants and top BLASTN hits from NCBI. Bootstrap support values from 1000 rapid bootstrap replicates are shown above branches. The name of each candidate gene copy is in bold adjacent to its respective tree. Mitogenomic reference sequences (and BLASTN hits to mitogenomes) are indicated. Sequences from Cuscuta are highlighted in blue. Scale bars indicate inferred nucleotide substitutions per site.

Additional file 3: Fig. S14. Phylogenetic trees for mitochondrial genes across angiosperms, with particular focus (red) on Cuscuta species. Bootstrap support > 60% is shown adjacent to nodes. Scale bars indicate inferred nucleotide substitutions per site. Groups are coloured according to the legend below, with the exception of Cuscuta (red). The names of the genes are indicated above their respective trees.

Additional file 4: Table S3. Reference plastomes used for annotation and BLASTN searches and their GenBank accession numbers. Table S4. Reference mitogenomes used for annotation, BLASTN searches and phylogenetic analysis and their GenBank accession numbers.

Acknowledgements
Not applicable.

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.

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]. The funding bodies played no role in the design of the study; the collection, analysis, and interpretation of data; and the writing of the manuscript. Open Access funding provided by Stockholm University.

Availability of data and materials
Organelle genomes are deposited in GenBank under accessions BK059221 and BK059222 for the C. australis and C. campestris plastomes, and BK059197-BK059204 and BK016277 for the C. australis and C. campestris mitogenomes, respectively. Accession numbers for reference sequences downloaded from Genbank are provided in Additional file 4. Sequence alignments used in phylogenetic analysis are included as supplementary material for an earlier version of this paper at https://doi.org/10.1101/2021.04.15.439983.

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
1Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden. 2Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Breivika, 9037 Tromsø, Norway.

Received: 21 September 2020 Accepted: 19 October 2021
Published online: 12 November 2021

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. Ounning LT, Oforsen JK, Parsiod 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, Wafuka 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. https://doi.org/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, Wafuka EK, de Pamphilis CW, Westwood JH. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. Science. 2014;345:808–11.
7. Rice DW, Alversen AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, et al. Horizontal transfer of entire genes 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. Berghorston 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. Koell宁chenko M, Konstantinov Y, Dietrich A. Plant mitochondria actively import DNA via the permeability transition pore complex. EMBO J. 2003;22:1245–54.
12. Logan DC. Plant mitochondrial dynamics. Biochim Biophys Acta. 1763;2018:430–41.
13. Arita Y, Yamamoto J, Aida GP, Nakazono M, Tsutsui N. Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. FEBS Lett. 2006;557:66–73.
14. Davis CC, Xi Z. Horizontal gene transfer in parasitic plants. Curr Opin Plant Biol. 2015;26:14–9.
15. Petersen G, Anderson B, Braun H-P, Meyer EH, Møller IM. Mitochondria in parasitic plants. Mitochondron. 2020;52:173–82.
16. 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.
17. Krause K. From chloroplasts to “cryptic” plastids: evolution of plastid genomes in parasitic plants. Curr Genet. 2008;54:111.
18. Petersen G, Cuenca A, Möller IM, Seberg O. Massive gene loss in mistletoe (Viscum, Viscaceae) mitochondria. Sci Rep. 2015;5:17588.
19. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant Viscum scurulideum is extremely divergent and dynamic and has lost all rna genes. Proc Natl Acad Sci U S A. 2015;112: E3515–24.
20. Zervas A, Petersen G, Seberg O. Mitochondrial genome evolution in parasitic plants. BMC Evol Biol. 2019;19:187.
21. Fan W, Zhu A, Kozaczek M, Shah N, Pabón-Mora N, González F, et al. Limited mitogenomic degradation in response to a parasitic lifestyle in Orobanchaceae. Sci Rep. 2016;6:36285.
22. Sanchez-Puerta MV, Ederer 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 (Balanophonaceae). Mol Phylogenet Evol. 2019;132:243–50.
23. Sanchez-Puerta M, Garcia LE, Edera AA, Palmer JD, Sato H, Sanchez-Puerta MV. Horizontal gene transfer in a parasitic flowering plant clade. PLoS Genet. 2013;9:e1003265.
24. Garcia LE, Ederer AA, Palmer JD, Sato H, Sanchez-Puerta MV. Horizontal gene transfer dominates the functional mitochondrial gene space of a holoparasitic flowering plant. New Phytol. 2020; https://doi.org/10.1111/nph.17926.
25. Dawson JH, Musselman LJ, Wolswinkel P, Dör I. Biology and control of Cuscuta. Rev Weed Sci. 1994;6:265–317.
26. van der Kooij TA, Krause K, Dör I, Krupinska K. Molecular, functional and ultrastructural characterisation of plastids from six species of the parasitic flowering plant genus Cuscuta. Planta. 2000;202:101–7.
27. McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. Complete plastid genome sequence of Geranium asymmetricum (Geraniaceae). Nucleic Acids Res. 2005;33:7053–62.
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, Olimståeg RD. 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 plant Cuscuta. J Exp Bot. 2005;56:2477–86.
31. Bischwillks M, Haapt S, Hofius D, Neumann S. Transfer of phloem-mobile substances from the host plants to the holoparasite Cuscuta sp. J Exp Bot. 2006;57:911–21.
32. Vogel A, Schwacke R, Denton AK, Uaadel 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.
33. Mower JP, Stefanović S, Young GJ, Palmer JD. Plant genetics: gene transfer from parasitic to host plants. Nature. 2004;432:165–6.
34. 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.
35. Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, et al. Dynamic evolution of Gcn4 mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol. 2015;208:570–83.
36. Jiang L, Wijeratne AJ, Wijeratne S, Fraga M, Meulia T, Doohan D, et al. Profiling mRNAs of two Cuscuta species reveal possible candidate transcripts shared by parasitic plants. PLoS One. 2013;8:e81389.
37. Zhang L, 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 genus Mimulus. Mol Biol Evol. 2019;36:154–68.
38. Zhang L, Kuo AC, Zhong Y, Lai H, Luo S, Cui F, et al. Identification of a trans-splicing intron of the nuclear ITGAM gene in diverse plant species. Nat Commun. 2018;9:2683.
39. Krause K, Dietmann CL. Analysis of transcription asymmetries along the trRNA-CCG operon; evidence for transcription attenuation and rapid RNA degradation between coding sequences. Nucleic Acids Res. 2004;32:276–83.
40. Zhang Y, Fernandez-Aparicio M, Wafula EK, Das M, Jiao Y, Wickett NJ, et al. Large-scale gene losses underlie the genome evolution of parasitic plant Cuscuta australis. Nat Commun. 2018;9:2683.
65. Choi IS, Ruhlman TA, Jansen RK. Comparative mitogenome analysis of the genus Trifolium reveals independent gene fission of ccnF and intracellular gene transfers in Fabaceae. Int J Mol Sci. 2020;21:1959.
66. Andrews S. FastQC. Babraham bioinformatics. 2018. https://www.bioinforma-
tics.babraham.ac.uk/projects/fastqc/. Accessed Feb 2019.
67. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory. 2014. https://sourceforge.net/projects/bbmap/.
Accessed 2 Oct 2018.
68. 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;27:72–36.
69. Carpenter EJ, Mataric N, Aylampalayam S, Wu S, Sun J, Yu J, et al. Access to RNA-sequencing data from 1,173 plant species: the 1000 plant transcriptome initiative (1KP). Gigascience. 2019;8. https://doi.org/10.1093/ gigascience/giz126.
70. Wei C, Tao X, Li M, He B, Yan L, Tan X, et al. De novo transcriptome assembly of Ipomoea nil using Illumina sequencing for gene discovery and SSR marker identification. Mol Gen Genomics. 2015;290:1873–84.
71. 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.
72. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
73. Altschul SF, Madden TL, Schaffer 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.
74. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
75. 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.
76. Wick RR, Judd LM, Gorrie GL, Holt KE. Univircycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595.
77. 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.
78. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.
79. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016;32:103–10.
80. Vaser R, Sovič I, Nagarajan N, Sikč M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737–46.
81. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45:e18.
82. 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.
83. Katoch K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
84. Robinson JT, Thorvaldsdóttir H, Winckler W, Gutmann M, Lander ES, Getz G, et al. Integrative genomic viewer. Nat Biotechnol. 2011;29:24–6.
85. Alverston 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. 2012;29:1436–48.
86. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.
87. Lowe TM, Eddy SR. T-RNAScan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:956–64.
88. Darzentas N. Circoletto: visualizing sequence similarity with Circos. Bioinformatics. 2010;26:2620–1.
89. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
90. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1001188.
91. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91.