2019

Plastome-wide rearrangements and gene losses in carnivorous Droseraceae

Paul G. Nevill

Katharine A. Howell
_The University of Notre Dame Australia_, kate.howell@nd.edu.au

Adam T. Cross

Anna V. Williams

Xiao Zhong

See next page for additional authors

Follow this and additional works at: https://researchonline.nd.edu.au/health_article

Part of the Life Sciences Commons, and the Medicine and Health Sciences Commons

This article was originally published as:

Original article available here:
10.1093/gbe/evz005

This article is posted on ResearchOnline@ND at
https://researchonline.nd.edu.au/health_article/256. For more information, please contact researchonline@nd.edu.au.
Authors
Paul G. Nevill, Katharine A. Howell, Adam T. Cross, Anna V. Williams, Xiao Zhong, Julian Tonti-Filippini, Laura M. Boykin, Kingsley W. Dixon, and Ian Small
Plastome-Wide Rearrangements and Gene Losses in Carnivorous Droseraceae

Paul G. Nevill¹²³, Katharine A. Howell⁴⁶, Adam T. Cross¹²³, Anna V. Williams²⁴, Xiao Zhong⁴⁵, Julian Tonti-Filippini⁴⁵, Laura M. Boykin⁴⁵, Kingsley W. Dixon¹² and Ian Small⁴⁵,*

¹ARC Centre for Mine Site Restoration, School of Molecular and Life Sciences, Curtin University, Bentley, Western Australia, Australia
²School of Plant Biology, The University of Western Australia, Crawley, Western Australia, Australia
³Kings Park and Botanic Garden, Kings Park, Western Australia, Australia
⁴Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, Western Australia, Australia
⁵School of Molecular Sciences, The University of Western Australia, Crawley, Western Australia, Australia
⁶Present address: The University of Notre Dame, Fremantle, Western Australia, Australia

*Corresponding author: E-mail: ian.small@uwa.edu.au.
Accepted: January 5, 2019

Data deposition: The full sets of unassembled reads are available at SRA under the accessions SRR7072322–SRR7072325 and SRR7072766–SRR7072769. The assembled, annotated sequences are available at GenBank under the accessions KY651214 (Drosera erythrorhiza), KY679199 (Drosera regia), KY679200 (Aldrovanda vesiculosa), and KY679201 (Dionaea muscipula).

Abstract
The plastid genomes of four related carnivorous plants (Drosera regia, Drosera erythrorhiza, Aldrovanda vesiculosa, and Dionaea muscipula) were sequenced to examine changes potentially induced by the transition to carnivory. The plastid genomes of the Droseraceae show multiple rearrangements, gene losses, and large expansions or contractions of the inverted repeat. All the ndh genes are lost or nonfunctional, as well as in some of the species, clpP1, ycf1, ycf2 and some tRNA genes. Uniquely, among land plants, the trnK gene has no intron. Carnivory in the Droseraceae coincides with changes in plastid gene content similar to those induced by parasitism and mycoheterotrophy, suggesting parallel changes in chloroplast function due to the similar switch from autotrophy to (mixo-) heterotrophy. A molecular phylogeny of the taxa based on all shared plastid genes indicates that the “snap-traps” of Aldrovanda and Dionaea have a common origin.

Key words: carnivorous plants, chloroplast genome, Droseraceae, gene loss, intron loss.

Introduction
Carnivorous plants derive a range of amino acids, peptides, macro-, and micro-nutrients from captured prey, and studies on numerous species indicate that significant resources are invested in the production of trapping organs (Plummer and Kethley 1964; Dixon et al. 1980; Adamec 1997; Adlassnig et al. 2012; Adamec 2013; Fasbender et al. 2017). Carnivory is most successful in open habitats on nutrient-poor soils and in oligotrophic aquatic environments, where it is likely to provide a competitive advantage (Givnish et al. 1984; Pavlović and Saganová 2015). It has arisen independently at least nine times in five different orders (Caryophyllales, Ericales, Lamiales, Oxalidales, and Poales) (Albert et al. 1992; Givnish 2015), and around 700 species are currently considered carnivorous (Ellison and Adamec 2017).

Droseraceae is the largest family of carnivorous plants, comprising three extant genera with distinct morphologies: Drosera, Dionaea, and Aldrovanda. Although Aldrovanda and Dionaea are both monotypic, Drosera rivals Utricularia of the Lentibulariaceae as the largest carnivorous genus, with around 250 members (Gonella et al. 2016; Fleischmann et al. 2017). Members of the cosmopolitan Drosera vary extensively in their specific morphology, but all have highly modified leaves lined with tentacle-like glandular trichomes that secrete a sticky substance to ensnare and digest prey.
Tentacle-bending and enzyme secretion are triggered by prey movements and chemical cues that induce action potentials in the contacting tentacles and subsequently local accumulation of jasmonates (Krusko et al. 2017). Dionaea and Aldrovanda also possess highly modified leaves, but unlike Drosera have snap-traps that close rapidly upon excitation (Poppinga et al. 2013). Dionaea muscipula is terrestrial and is restricted to a small area of the southeastern United States (Bailey and McPherson 2014), whereas Aldrovanda vesiculosa is a submerged free-floating aquatic found in Europe, Asia, Australia, and Africa (Cross 2012).

At the molecular level, studies of genome evolution in carnivorous lineages have found dramatic differences when compared with solely autotrophic plant species (Fukushima et al. 2017). For example, carnivorous Lentibulariaceae exhibit the smallest angiosperm nuclear genomes (Ibarra-Laclette et al. 2017). This genome instability contrasts with the general observation that plastid genomes are highly conserved in all genetic compartments (Jobson and Albert 2002). The plastid genome is also reduced in size, mostly due to the independent loss of genes for the plastid NAD(P)H dehydrogenase (Wicke et al. 2014). There are also altered proportions of plastid repeat DNA; a significant plastome-wide increase of substitution rates/microstructural changes; a disproportionate elevation of nonsynonymous substitutions, particularly in protein-coding genes; and shifts in selective regimes relative to noncarnivores (Wicke et al. 2014). This genome instability contrasts with the general observation that plastid genomes are highly conserved in sequence, gene content, and gene order in flowering plants (Green 2011; Wicke et al. 2014). Unusual modes of plastid genome and plastid gene evolution are rare in autotrophic lineages (Palmer et al. 1988; Cosner et al. 2004; Sloan et al. 2012). Plastome instability and gene loss appears to be associated with the switch from an autotrophy to (mixo-) heterotrophy. Elevated plastome-wide substitution rates and the loss of ndh genes observed in Lentibulariaceae were also found in obligate photosynthetic, parasitic plants including Orobancheaeae (Wicke et al. 2013; Frailey et al. 2018) and Cuscuta (McNeal, Arumugunathan, et al. 2007; McNeal, Kuehl, et al. 2007) and in mycoherbotoerths (Graham et al. 2017). This raises a question: Is the rapid evolution of plastid genomes promoted by different nutrient acquisition strategies?

Advances in high-throughput sequencing have greatly increased the ability to characterize whole genomes, particularly from organelles (Tonti-Filippini et al. 2017). Combined with analytical tools and databases that facilitate genome annotation and comparisons with previously sequenced plastid genomes, studies of genome evolution in nonmodel organisms are increasingly accessible. To explore the effect of the carnivorous syndrome on the evolution of the chloroplast, we sequenced the chloroplast genomes of the three extant genera (Drosera, Dionaea, and Aldrovanda) of the carnivorous plant family Droseraceae. The results of this study improve our understanding of the genomic effects of a carnivorous strategy and provide further evidence of the congruence between carnivorous and parasitic plastid genomes.

Materials and Methods

Plant Material and DNA Extraction

Plant material of A. vesiculosa L. was collected from Cape le Grande, Esperance, Western Australia (voucher: PERTH 08722560) (Cross 2012). Plant material of Drosera erythrorhiza Lindl. was collected from Alison Baird Reserve, Kenwick, Western Australia (voucher: PERTH 06332374). Plant material of D. muscipula Sol. ex J.Ellis and Drosera regia Stephens were collected from cultivated individuals maintained in the living collection at Kings Park and Botanic Gardens, Perth, Western Australia. Total genomic DNA was extracted from fresh material of all four species using a cetyl trimethylammonium bromide extraction method (Doyle and Doyle 1987). Cetyl trimethylammonium bromide extraction buffer (1 ml) was added to each sample in a 2-ml tube containing a scoop of sterilized silica sand and two small glass beads. Tissue was macerated twice for 30 s at speed four (6.5 m/s) in a Thermo Savant FASTPREP FP120. All samples were treated with RNase (Qiagen), and DNA quality and quantity were assessed using a NanoDrop spectrophotometer (ND-1000; Thermo Fisher Scientific, USA), a Qubit 2.0 fluorometer (Invitrogen, Life Technologies) and via agarose gel electrophoresis. DNA samples were purified and concentrated using a DNA Clean & Concentrator-5 kit (Zymo Research).

Illumina Sequencing

Total genomic DNA (30–150 ng) was fragmented with a Covaris S2 to a mean fragment size of 550 bp. DNA libraries were prepared using the TruSeq DNA Nano Library Prep kit (Illumina), according to the manufacturer’s instructions. Libraries were quantified by quantitative polymerase chain reaction, using a KAPA Illumina Library Quantification kit (KAPA Biosystems) and LightCycler 480 (Roche), and then pooled in approximately equimolar amounts. The pooled libraries were sequenced (2 × 150 bp) using the Illumina HiSeq 1500 in rapid run mode using a TruSeq Rapid PE Cluster kit, for on-board cluster generation, and TruSeq Rapid SBS kits.

Sequence Assembly and Annotation

For each specimen, overlapping paired-end reads were merged using the software FLASH v. 1.2.7 (Magoc and Salzberg 2011) to maximize the effective read length. Merged reads were assembled de novo using Velvet v. 1.2.08 (Zerbino and Birney 2008) with k-mer values of between 71 and 111 and the coverage cutoff optimized for each sample and k-mer value to reduce contamination by nuclear and mitochondrial reads. For D. muscipula, A.
vesiculosa, and Dr. regia, complete contigs corresponding to the large and small single copy regions and the inverted repeat were obtained from the initial assembly. For Dr. erythrorhiza, five contigs were obtained (table 1). For each assembly, MUMmer v. 3.0 (Kurtz et al. 2004) was used to compare the assembled chloroplast contigs with other plastid genomes, particularly that of *Fagopyrum esculentum* (RefSeq accession NC_010776) to order the contigs correctly. The assemblies were refined by mapping the original reads to the draft genomes with bowtie2 (using default parameters) (Langmead and Salzberg 2012), verifying the assembly with genomes with bowtie2 (using default parameters) were refined by mapping the original reads to the draft NC_010776) to order the contigs correctly. The assemblies were also attempted using two seed-extension assemblers, The Organelle Assembler v. b0 attempted using two seed-extension assemblers, The Organelle Assembler v. b2.2’ (https://git.metabarcoding.org/org-asm; last accessed February 3, 2019, parameters “--minread 5 --smallbranches 15 --seeds protChloroArabidopsis”) and NOVOPlasty v. 2.5.9, (using k-mer length 39 and the Zea mays rbcL gene as a seed) (Dierckxsens et al. 2017). The resulting single-contig assemblies were permuted to match canonical plastid genome arrangements by flipping the single copy regions if necessary. The genomes were initially annotated by the “Transfer Annotations” function within Geneious 9.1.5 (Kearse et al. 2012) (using *F. esculentum* as a reference at a 90% identity threshold) followed by manual curation using similarity to each other and to other previously annotated genomes as a guide. Genes were only annotated as present when there was a reasonable expectation that the gene product might be functional, that is, degenerate sequences missing conserved domains, or containing frameshifts or internal stop codons were not annotated. The full sets of unassembled reads are available as SRA accessions SRR7072322–SRR7072325 and SRR7072766–SRR7072769. The assembled, annotated sequences are available as GenBank accessions KY651214 (Dr. erythrorhiza), KY679199 (Dr. regia), KY679200 (A. vesiculosa), and KY679201 (D. muscipula).

### Verification of Gene Losses

Reads from samples with apparently missing plastid genes were mapped to related genomes containing the gene to verify if reads matching the gene sequence were present in the sample, but not assembled into the final genome. This was carried out using bowtie2 (Langmead and Salzberg 2012) (default parameters) and bwa v. 0.7.15 (bwa-mem algorithm, -T 25) (Li and Durbin 2009). Cross mappings of this type were done for *Dr. erythrorhiza* against *Drosera rotundifolia* and *Dr. regia*, and for all of the Droseraceae sequenced here against the *F. esculentum* genome. In no cases were additional plastid sequences found that had not been included in the assemblies. Contig pools generated during the assemblies were also checked by TBlastN (with an e-value cutoff of $10^{-5}$) using the protein sequences of CipP1, Ycf1, Ycf2, Rpl32, and NdhA as queries. No contigs not already included in the assemblies were discovered.

### Phylogenetics

Fifty-eight proteins are encoded in each of the five Droseraceae plastid genomes and that of *F. esculentum*. Each set of six protein-coding (i.e., nucleotide) sequences was aligned independently using TranslatorX v. 1.1 (Abascal et al. 2010) to align codons with MAFFT v. 7.407 (Katoh and Standley 2013) as the alignment engine. The alignments were then trimmed with trimAl v1.4 (–automated1) (Capella-Gutierrez et al. 2009) and concatenated. Bayesian analyses were conducted using MrBayes v. pre-3.2.7 (Ronquist et al. 2012). The aln.model file consisted of 58 partitions corresponding to each protein-coding sequence and the parameters were estimated based on these partitions. The alignment and the resulting MrBayes phylogeny are available from TreeBASE (study TB2: S23621). Thirty independent analyses were run simultaneously for 30 million generations with sampling every 300 generations. Each analysis consisted of two independent runs, each utilizing four coupled Markov chains. The run convergence was monitored by finding the plateau in the likelihood scores (standard deviation of split frequencies <0.0015). The first 25% of topologies were discarded as burn-in for the estimation of a majority rule consensus topology and posterior probability for each node on the phylogenetic tree. Finally, Tracer v. 1.7.1 (Rambaut et al. 2018) was used for analyzing the trace files generated by the Bayesian MCMC runs to identify potential convergence problems. The same alignments were used to generate maximum-likelihood trees with RAxML v. 8.2.12 (Stamatakis 2014), using -m GTRGAMMAI and 1,000 rapid bootstrap inferences, and with IQ-TREE v.1.6.8 (Nguyen et al. 2015) using automatic selection of the best-fit model (Kalyaanamoorthy et al. 2017), with -a lrt 100000 -bb 100000.

### Results

#### Structure of the Plastid Genomes in Droseraceae

Illumina sequencing of libraries prepared from total DNA of *D. muscipula* (Venus’ fly trap), *A. vesiculosa* (waterwheel plant), *Dr. regia* (king sundew), and *Dr. erythrorhiza* (red ink sundew) produced between 5 and 64 million paired-end reads per sample with a length of 150 nt. For each specimen, ~5% of the reads could be assembled into contigs homologous to the reference plastid genome of *F. esculentum* (Logacheva et al. 2008), the most closely related noncarnivorous plant whose plastid genome has been sequenced. Read coverage was relatively even across the genomes and...
exceeded 400 in all cases (table 1). Complete assemblies were obtained for all four species, ranging from 117,589 nt (D. muscipula) to 141,568 nt (A. vesiculosa) (fig. 1). Similar assemblies were obtained with three different assemblers (table 1). Minor unresolved discrepancies between assemblies were largely due to differences in the copy number of short tandem repeats in low complexity intergenic regions; these variations were present in the sequencing reads and may be present in vivo. These genomes are considerably smaller than the reported 192,912 nt size of the D. rotundifolia genome (RefSeq accession NC_029770). The differences in genome sizes are primarily due to large differences in the extent of the inverted repeats and secondarily due to differences in gene content.

**Expansion and Contraction of the Inverted Repeats**

Based on the generally typical genome size and arrangement of *F. esculentum*, the ancestral plastid genome of the Droseraceae is likely to have contained a typical inverted

<table>
<thead>
<tr>
<th>Genome Size (bp)</th>
<th>velvet</th>
<th>Organelle Assembler</th>
<th>NOVOPlasty</th>
<th>IR Length (bp)</th>
<th>Mapped Reads</th>
<th>Mean Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Dionaea muscipula</em></td>
<td>117,589</td>
<td>57,689</td>
<td>117,692</td>
<td>91,604</td>
<td>2,798</td>
<td>1,313,104</td>
</tr>
<tr>
<td><em>Aldrovanda vesiculosa</em></td>
<td>141,568</td>
<td>75,654</td>
<td>141,583</td>
<td>141,573</td>
<td>27,040</td>
<td>484,787</td>
</tr>
<tr>
<td><em>Drosera regia</em></td>
<td>136,810</td>
<td>79,603</td>
<td>136,810</td>
<td>136,810</td>
<td>23,586</td>
<td>827,359</td>
</tr>
<tr>
<td><em>Drosera erythrorhiza</em></td>
<td>134,391</td>
<td>51,587</td>
<td>129,708</td>
<td>54,762</td>
<td>37,186</td>
<td>425,913</td>
</tr>
</tbody>
</table>

*Note.* *-_* indicates contigs that were duplicated in the final assembly as part of the inverted repeats.
repeat of 25–30 kb containing the \textit{rrn} genes, several tRNA genes (including \textit{ttnl-GAU} and \textit{ttnA-UGC}), several genes encoding ribosomal subunits, \textit{ndhB}, and most or all of the \textit{ycf1} and \textit{ycf2} coding sequences. All of the Droseraceae have a modified IR region compared with \textit{F. esculentum}, with an extreme variation in size from under 3 kb in \textit{D. muscipula} to almost 40 kb in \textit{Dr. erythrorhiza} and \textit{Dr. rotundifolia} (table 1 and fig. 1), with concomitant variation in gene content from only 1 gene entirely within the IR in \textit{D. muscipula} (\textit{rrn16}) up to 40 genes contained within the IR of \textit{Dr. erythrorhiza}. These variations are not simply movements of the IR borders but include multiple gene losses and other rearrangements resulting in a lack of colinearity throughout the IR region between these genomes. The very unusual size and structure of the IR in \textit{D. muscipula} was confirmed by polymerase chain reaction (supplementary fig. S1 and supplementary table S1, Supplementary Material online).

### Gene Content of the Plastid Genomes in Droseraceae

The \textit{F. esculentum} plastid genome (Logacheva et al. 2008) includes 114 unique genes, of which 11–21 are missing from the Droseraceae plastid genomes (fig. 2). The 11 plastid genes encoding subunits of the ferredoxin-plastoquinone oxidoreductase (NDH) complex are missing (or rearranged to the point that they cannot possibly be functional) from all of the Droseraceae plastid genomes. Multiple additional gene losses have occurred within the Droseraceae. Of particular interest are the loss/inactivation of the usually essential protein-encoding genes \textit{clpP1} (supplementary fig. S2, Supplementary Material online), \textit{ycf1}, and \textit{ycf2} from \textit{Dr. erythrorhiza} and \textit{Dr. rotundifolia} and the loss of several equally essential tRNA genes: \textit{ttnA-UGC} from \textit{Dr. erythrorhiza} and \textit{Dr. rotundifolia}, \textit{tmV-UAC} from \textit{Dr. rotundifolia} and \textit{D. muscipula}, and \textit{ttnl-GAU} from \textit{Dr. erythrorhiza}. Gene losses were verified by mapping raw reads from the relevant sample against the nearest related genome containing the gene to make sure that apparent losses were not simply assembly errors. An example is shown in figure 3. In addition to complete gene losses, other genes have accumulated mutations, indels, and rearrangements to the point where their functionality in at least some of the species is doubtful. Examples include \textit{accD} and \textit{tnG-UCC} (fig. 4).

### Intron Losses

In addition to losses of entire genes, there are also losses of specific introns. Both introns are missing from the \textit{clpP1} genes of all the Droseraceae. These introns are also missing from several other plant lineages, including grasses, pines, and some dicot genera (Erixon and Oxelman 2008; Jansen et al. 2008). The intron of \textit{rpl2} has also been lost independently from several groups of flowering plants and was reported to be missing from \textit{Drosera filiformis} based on filter hybridization (Downie et al. 1991). We confirm the loss of this intron.

![Fig. 2.—Gene content in Droseraceae plastid genomes and in \textit{Fagopyrum esculentum}. Genes present and apparently functional are shaded in black, doubtfully functional genes in mid-gray, almost certainly nonfunctional remnants in pale gray and missing genes in white.](https://academic.oup.com/gbe/article-abstract/11/2/472/5284917/10.1093/gbe/evz005 Advance Access publication January 10, 2019)
from all the Droseraceae. A fourth intron missing from all of the plastid genomes we have sequenced is the second (cis-spliced) intron of rps12. Again, this is not unique among flowering plants as the same intron has been lost from legumes (Jansen et al. 2008) and from parasitic plants such as Cuscuta (McNeal et al. 2009). Splicing of rpl2 and rps12 intron 2 is thought to require the intron maturase MatK (Zoschke et al. 2010), so it is interesting to note that the arrangement of the trnK and matK genes in the Droseraceae is unique. In all other land plant genomes where it is present, the trnK gene contains a long intron in which lies the matK gene. In all the Droseraceae plastid genomes, the trnK gene lacks an intron, and in all but Dr. erythrorhiza, lies near to the matK gene, adjacent to the sequence corresponding to the 5′ tmK exon in other plant plastid genomes (fig. 5a). The sequence of the Droseraceae tmK differs from that in F. esculentum and other flowering plants (fig. 5b).

Changes in Gene Order
There are many differences in gene order and orientation between the different plastid genomes of the Droseraceae (fig. 1). Some of these rearrangements have affected the 18 cotranscribed gene sets whose order is otherwise highly conserved across flowering plants (Sugita and Sugiuura 1996) (supplementary fig. S3, Supplementary Material online). A summary is shown in table 2. Some of the alterations to transcription units are due to loss of one or more of the genes normally contained within, for example, loss of the ndh genes disrupts three commonly conserved transcription units. Others are due to rearrangements that break the linkage between adjacent genes but without the loss of any genes, for example, the separation of the rps11-rpoA gene pair from the rpl23…rpl36 gene cluster in Aldrovanda. As these rearrangements are unlikely to be reversible, they may be phylogenetically informative where events are shared between taxa.

Phylogenetic Analyses
The coding sequences shared by the Droseraceae plastid genomes were used to reconstruct their phylogenetic relationships, using the Fagopyrum genome as an outgroup (fig. 6). The tree strongly supports previous phylogenetic hypotheses regarding the relationships between Drosera, Dionaea, and Aldrovanda, with Aldrovanda + Dionaea forming a sister group to Drosera (Cameron et al. 2002; Rivadavia et al. 2003).

Discussion
Confirmation That “Snap-Traps” Evolved Only Once in the Droseraceae
The Droseraceae diverged relatively early from the other non-carnivorous Caryophyllales (around 65 Ma; Heubl et al. 2006), and the age of paleobotanical finds for ancestral Aldrovanda species (55–38 Ma) indicate that divergence within the family also occurred relatively early (Cross 2012). There has been much discussion about the evolution of the remarkable traps of the Droseraceae and the relationships between the diverse methods for capturing live prey that are found in the three genera (Poppinga et al. 2013). In particular, it has often been questioned whether the highly distinctive active “snap-traps” found in Dionaea and Aldrovanda have a common origin or evolved independently. Phylogenetic studies utilizing nuclear markers have placed Aldrovanda sister to Dionaea within Droseraceae (bootstrap support of around 90%), supported
by combined analyses of rbcL and nuclear 18S rDNA (Rivadavia et al. 2003), and rbcL, matK, atpB, and 18S rDNA (Cameron et al. 2002), as well as similarities in trichome and gland morphology (Elansary et al. 2010). Phylogenies based only on plastid DNA have either lacked sufficient resolution or placed Aldrovanda closer to members of Drosera (Cameron et al. 2002; Rivadavia et al. 2003), although this discrepancy has been attributed to chloroplast capture because of the incongruence between trees based on plastid genomes and nuclear sequences (Elansary et al. 2010). The molecular evidence for these relationships is now much stronger with the entire plastid genomes to analyze. Our phylogenetic analysis supported the tentative conclusions of the previous molecular and morphological studies, confirming the sister relationship between Aldrovanda and Dionaea (Elansary et al. 2010; Poppinga et al. 2013) and confirming that they form a clade that is sister to Drosera with the South African species Dr. regia sister to other Drosera sp. (Rivadavia et al. 2003). Our trees based on plastid sequences agree with the previous nuclear DNA phylogenies. Thus, our results do not support chloroplast capture and we suspect that the difficulty in establishing the correct phylogeny from plastid sequences in the past may be partly due to the large variations in genome structure and gene content within the Droseraceae. In particular, sequences within the inverted repeats tend to show a much reduced rate of fixation of mutations (Wolfe et al. 1987), and the extreme variation of the extent of these repeats within the Droseraceae may lead to differences in the rate of divergence of different genes in different lineages. For example, much of the rbcL gene, used
...for most of the molecular phylogenetics work to date (Cameron et al. 2002; Rivadavia et al. 2003), is within the IR in Dr. erythrorhiza but not in the other species studied. Similarly, we have shown that the matK gene, another popular choice for phylogenetic reconstructions, is in a different genomic environment in the Droseraceae than in other flowering plants, with no doubt a different set of selection pressures acting on it and its flanking sequences.

Further evidence for the relationship and origin of the Droseraceae can be found in the pattern of gene function losses. The unique loss of the trnK intron proves a common origin for the three genera and thus confirms the monophyly of the Droseraceae. There are no losses specific to the Aldrovanda + Dionaea clade, but the disruption of the petL...rps18 transcription unit between rpl33 and rps18 is only seen in these two taxa and is probably a synapomorphy. The position of Dr. regia as sister to the rest of the Drosera clade is confirmed by the losses of trnA and ycf1 and several other rearrangements that are shared between the northern hemisphere plant Dr. rotundifolia and the southern hemisphere Dr. erythrorhiza but not with Dr. regia. Drosera regia is restricted to a single mountain valley in South Africa and has several pleomorphic characters (e.g., operculate pollen, a lack of stipules) that are more similar to those in D. muscipula, and which traditionally has resulted in being treated as a different group to other Drosera (Fleischmann et al. 2017).

An Unusual trnK Gene Suggests an Unlikely Event

The original trnK sequence and associated intron is still present (if barely recognizable) in all the Droseraceae except Dr. erythrorhiza, where the 5′ exon and associated intron is missing (fig. 5). This trnK pseudogene is clearly not functional as it is not conserved within the Droseraceae. Instead, there has been an insertion of a new trnK gene, which does not contain an intron, and which is conserved within the Droseraceae and thus presumably functional. In all the Droseraceae besides Dr. erythrorhiza, this new trnK gene is adjacent to the remains of the previous one, and this is presumably the original site of the insertion. In Dr. erythrorhiza, the new trnK gene is tens of kilobases away near the trnD gene. The loss of the trnK intron, without concomitant loss of the trnK gene, is unique (as far as we are aware) within the terrestrial flora. Insertion of the matK-containing intron within trnK predates the colonization of the land by plants, as this intron is found throughout the charophyte clade (Hausner et al. 2006), except in a few plants (notably most ferns and many parasitic plants) that have lost trnK entirely (Duffy et al. 2009; McNeal et al. 2009).

We propose two possible explanations for the origin of the novel trnK gene in the Droseraceae. Both of them appear unlikely, but other explanations would seem even less likely. One possibility is that a spliced trnK transcript was reverse transcribed and reinserted into the genome. This is thought to be the mechanism by which intron losses in other plastid...
<table>
<thead>
<tr>
<th>Transcription Unit</th>
<th>Fagopyrum</th>
<th>Aldrovanda</th>
<th>Dionaea</th>
<th>Drosera regia</th>
<th>Drosera erythrorhiza</th>
<th>Drosera rotundifolia</th>
</tr>
</thead>
<tbody>
<tr>
<td>atpB-atpE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rps12-rps7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>psbE-psbF-psbL-psbL</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ndhD-ndhG-ndhH</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rrn16-tml-tmA-rm23-rrn4-5-rrn5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rpoB-rpoCl-rpoC2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-inrA-rpl36-rps11-rpoA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>trnE-tmY-trnD</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>clpP1-rps12-rp20</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>petL-petG-psaI-rpl33-rps18</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>psaA-psaB-rps14</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>psaC-ndhD</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>psbB-psbH-petB-petD</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>psbK-psbl-tmG</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ndhA-ndhI-ndhG-ndhE-psaC</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rpl32-ccsA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rps2-atpI-atpH-atpFatpA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
genes have occurred (Downie et al. 1991). However, the sequence of the Droseraceae trnK differs from that in other flowering plants. This hypothesis must therefore account for the fact that within a relatively short period of time (after the divergence from Fagopyrum but before the divergence within the Droseraceae), this sequence has accumulated several mutations not seen in other flowering plants. This may be possible if the requirement for intron splicing places a strong selective pressure on the trnK exons that is relaxed once splicing is no longer necessary. Alternatively, the Droseraceae trnK gene may be derived from horizontal transfer from another organism. Horizontal gene transfer between plant mitochondria has been documented on multiple occasions (Bergthorsson et al. 2003, 2004), and of functional tRNA genes from plastids to mitochondria (Joyce and Gray 1989). However, as far as we are aware, horizontal gene transfer has not been observed between plastids in land plants, although it has been shown in algae (Ruck et al. 2017). However, as the matK intron insertion into trnK occurred prior to the appearance of charophyte algae, if the transferred sequence was DNA, this implies the source would have to have been a non-charophyte alga. Algal plastid trnK sequences are reasonably similar (in some cases >90% identical) to the Droseraceae sequence, and as the Droseraceae are often associated with wet or aquatic habitats such a transfer is perhaps not entirely outside the realms of possibilities.

The matK gene is retained in the Droseraceae, despite the loss of the trnK intron. This is similar to the case in plants where matK is retained despite complete loss of trnK (Duffy et al. 2009; McNeal et al. 2009), and adds to the evidence that the MatK maturase is essential for splicing other plastid introns (Zoschke et al. 2010). Three of the introns thought to require MatK are missing from the Droseraceae plastid genomes (trnK, rpl2, and rps12 intron 2) but several putative MatK targets remain. There is an interesting parallel in Geraniaceae mitochondria, where loss of several nad1 introns has left the matR gene (usually encoded within nad1 intron 4) as a free-standing and presumably functional gene (Grewe et al. 2016), because it is required for splicing of introns in other genes (Sultan et al. 2016).

Functional Implications of Genome Rearrangements and Gene Losses

The main focus of this study was to explore the effect of the carnivorous syndrome on evolution of the plastid genome in Droseraceae and examine any resultant alteration in plastid function. All three extant genera in Droseraceae have highly rearranged plastid genomes, including some rarely or never before seen events. These include the losses of several genes usually thought to be essential in plant plastids. Only four protein-coding genes are consistently retained in most plastid genomes of nonphotosynthetic parasitic plants that have lost most of their plastid genes: accD, ycf1, ycf2, and clpP1 (Wolfe, Morden, and Palmer 1992; Delannoy et al. 2011; Graham et al. 2017). All four genes are essential when tested by gene knockout in tobacco (Drescher et al. 2000; Kuroda and Maliga 2003; Kode et al. 2005), and yet at least one of
them is missing or appears to be nonfunctional in each of the Droseraceae plastid genomes. Notably, Dr. rotundifolia and Dr. erythrorhiza appear to have lost the function of all four. Despite this rather startling set of plastid gene losses, the Droseraceae retain the ability to make fully functional chloroplasts, and indeed are highly successful at deriving carbon and energy from autotrophy even in the absence of prey capture (Bruzzone et al. 2010; Pavlović et al. 2014). It seems probable therefore that the accD (required for acetyl-CoA carboxylase) and ctpP1 (required for the Clp protease) genes are functionally replaced by nuclear counterparts, either because the plastid genes have been transferred or because nuclear homologs have gained plastid functions (exemplified by the nuclear ACC2 gene in Brassicaceae; Parker et al. 2016). The fate of ycf1 and ycf2 is less clear; in other plants, loss of ycf1 and ycf2 is not usually associated with transfer of these genes to the nucleus, and as the function of the gene product is not fully understood, what complements their loss is not known. Ycf1 has been proposed to be a key component of the general protein import channel (Kikuchi et al. 2013). However, this has been challenged (de Vries et al. 2015; Bölter and Soll 2017), one of the reasons being that multiple independent losses of this gene have occurred in plant plastids, which seems hard to reconcile with such an essential function. The Droseraceae provide an additional example that requires explanation. The ycf1 and ycf2 genes, when lost, are often lost in quick succession, strongly hinting that they act in the same pathway or function, and thus that the absence of both can be complemented by the same (unknown) process. There is also some indication that loss/reten-
tion of accD also correlates with that of ycf1/ycf2 (Delannoy et al. 2011; de Vries et al. 2015), a correlation that is strengthened by our analysis of the Droseraceae genomes. A role for Ycf1 in the assembly of ACCase has been proposed, based on this correlation ( Bölter and Soll 2017). These three genes are notoriously variable and difficult to annotate, such that it is difficult to be sure whether a particular plastid sequence is nonfunctional or simply particularly divergent (Nakai 2015), and indeed we are not entirely certain of the functional status of accD in the Droseraceae. A concerted effort to reanalyze the annotations of these genes across land plant plastid genomes would be difficult, but worthwhile.

Besides the loss of “essential” protein-coding genes, the Droseraceae plastid genomes are also missing several tRNA genes found in almost all other plastid genomes. For example, Dr. erythrorhiza is missing two essential tRNA genes (trnA-UGC and trnI-GAU) and D. muscipula is missing trnV-UAC. Loss of plastid tRNA genes is common in parasitic plants that have lost the ability to photosynthesize (Wolfe, Morden, Ems, et al. 1992; Delannoy et al. 2011) and thus no longer require high rates of protein translation in their plastids but is unusual in photosynthetic plants. As plastid translation must occur in the Droseraceae, we presume that the missing tRNAs can be imported from the cytosol. All three of the equivalent tRNAs can be imported into plant mitochondria (Kumar et al. 1996), so it may simply be a question of co-opting the requisite machinery, but the process is likely to limit the maximal translation rates possible in Droseraceae plastids (Salinas et al. 2012). The rearrangement of the rRNA operons may also have implications for plastid translation. The typical rrn16-trnl-trnA-rrn23-rrn4.5-rrn5 structure is highly conserved in plastids, and almost always within the inverted repeat, which helps preserve the sequence (Wolfe et al. 1987) and doubles the copy number of the rrn genes to permit high rates of ribosome synthesis. Within the Droseraceae, several alterations to this operon have occurred. Deletions between the rrn16 and rrn23 genes have removed part of trnA and trnl in Dr. rotundifolia and Dr. erythrorhiza, and severe contraction of the inverted repeats in D. ionaea have left the genome with only one complete copy of the rrn operon, plus a second copy of rrn16. These rearrange-
ments must have consequences for the processing of rRNA precursors and are likely to limit the maximal rate of plastid ribosome biogenesis.

Loss of ndh Genes in Carnivorous Plants

The genome rearrangements, gene losses, and general variability of the plastid genomes in the Droseraceae are reminiscent of what has been observed in other plants that are no longer entirely reliant on photosynthesis for energy and nutrients, such as hemiparasitic plants (McNeal, Kuehl, et al. 2007), partial mycoheterotrophs (e.g., orchids: Kim et al. 2015; Lin et al. 2017), and other carnivorous plants such as the Lentibulariaceae (Wicke et al. 2014). In particular, common to almost all of these plants is the early loss of the 11 ndh genes encoding subunits of the ferredoxin-plastoquinone oxido-reductase (NDH) complex (Ruhlman et al. 2015; Silva et al. 2016). This complex is related to mitochondrial complex I (NADH-ubiquinone oxidoreductase) and is thought to be able to transport protons across the thylakoid membrane in concert with cyclic electron flow around photosystem I, and in doing so can adjust the relative yield of ATP or reductant from photosynthetic electron flow (Shikanai 2016). The exact physiological role of the complex in most plants is still rather unclear, as mutants lacking it show few obvious abnormalities. It has variously been proposed to provide advantages under low light, high light, cold, or other “stress” conditions (Yamori et al. 2011). The loss of the NDH complex in partially parasitic, mycoheterotrophic, or carnivorous plants (and some other groups of plants with extensive mycorrhizal associations, such as the Pinaceae) may represent a loss of selection for optimal photosynthesis (Ruhlman et al. 2015), or it may be hinting at a specific function required only in fully autotrophic plants, presumably linked to nutrient assimilation. One candidate would be assimilation of nitrate, which requires photosynthesis-derived reductant.
Conclusion

The gene loss patterns observed here are similar to those in carnivorous *Utricularia*, *Pinguicula*, and *Genlisea* (Wicke et al. 2014; Silva et al. 2016). Finding comparable patterns in the phylogenetically distinct Droseraceae strongly supports convergent evolution in the plastid genome, presumably driven by the switch to a carnivorous lifestyle. The changes observed here in plastid gene content also resemble those found in parasitic (e.g., McNeal, Arumuganathan, et al. 2007; McNeal, Kuehl, et al. 2007; McNeal et al. 2009) and myco-heterotrophic plants (Graham et al. 2017), supporting a broader hypothesis of convergent plastid genome alterations in plants where there is a switch from photoautotrophy to a partially or fully heterotrophic mode of nutrition.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

This work was supported by the Australian Research Council (grant numbers IC150100041, FL140100179, and DE120101117). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. We thank Dan Sloan for comments and access to unpublished data on ClpP1 function.

Literature Cited


Associate editor: Daniel Sloan