Chloroplast genome evolution in New Zealand mycoheterotrophic Orchidaceae

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Plant Biology

at Massey University, Manawatu,
New Zealand

Katherine Jane Hope Murray

2019

Abstract

The plastid genomes, or plastomes, of most photosynthetic land plants are highly similar. In contrast, those of non-photosynthetic, heterotrophic land plants are often reduced in both size and gene content. The apparent degradation of mycoheterotrophic plant plastomes has been attributed to a functionally-driven stepwise pattern of loss. However, the number of complete plastome sequences available for mycoheterotrophic plants is small and taxonomic coverage is biased. In this thesis, the plastomes of two mycoheterotrophic orchid species endemic to New Zealand, *Corybas cryptanthus* Hatch (Diurideae) and *Danhatchia australis* Garay & Christenson (Goodyerinae), as well as those of an albino and several photosynthetic representatives of *Corybas* are reported. Beyond increasing the number of mycoheterotrophic plastomes available for evaluating broad hypotheses about plastome evolution in non-photosynthetic plants, these data also provide insights into two little studied aspects of plastome evolution in mycoheterotrophs; intraspecific variation in the plastomes of mycoheterotrophs and the differences between mycoheterotrophs and their closest photosynthetic relatives.

The plastomes of *C. cryptanthus* and *D. australis* differ in the extent to which they are degraded. Perhaps unexpectedly, the plastome of *C. cryptanthus*, which has close photosynthetic relatives and therefore is likely to have arisen more recently than the taxonomically isolated *D. australis*, is more reduced. Specifically, the plastomes of *C. cryptanthus* are approximately half the size and have half the gene content of the other *Corybas* sequenced whereas the plastome of *D. australis* is similar to those available for photosynthetic relatives. This contrast may reflect underlying differences between the two genera; the photosynthetic relatives of *D. australis* have plastomes containing NADH dehydrogenase (*ndh*) genes whereas those of photosynthetic *Corybas* have lost their *ndh* genes and their small single copy regions are highly reduced. These features may have predisposed the ancestor of *C. cryptanthus* to rapid genome degradation. Finally, observations on these results strongly suggest that plastome degradation follows, rather than precedes, the shift to mycoheterotrophy.

I

Acknowledgements

I would like to express my deep gratitude to my research supervisors, Dr Richard Winkworth, Dr Jennifer Tate and Dr Carlos Lehnebach, for their patient guidance, encouragement and feedback on this project. I would also like to thank Briana Nelson and Trish McLenachan for their support in the laboratory. My thanks are extended to Andrew Broome and Carl Timms, for helping me find albino and variegated *Corybas* specimens during field trips, and for their support and encouragement. In addition, I wish to thank Michael Moorhouse for providing photographs of *Corybas cryptanthus* and *Danhatchia australis*.

I would also like to extend my thanks to Graduate Women Manawatu (GWM) and Massey University for financial assistance, in the form of scholarships, to support my MSc. Likewise, the Museum of New Zealand Te Papa Tongarewa Herbarium (WELT) and Kew Gardens Herbarium are thanked for supplying DNA samples.

Contents

List of Figures

List of Tables

List of Abbreviations

accD	acetyl-coenzyme A carboxylase gene
ATP	adenosine triphosphate
atp	adenosine triphosphate (ATP) synthase subunit genes. Individual subunit genes are denoted A-F (i.e., *atpA-atpF*)
BLOSUM62	blocks substitution matrix 62
bs	bootstrap support
BWA	Burrows-Wheeler aligner
ccsA	cytochrome c biogenesis gene
cemA	chloroplast envelope membrane protein gene
clpP	ATP-dependent Clp protease gene
cm	centimetre
CoRR	co-location for redox regulation
DF	degrees of freedom
DNA	deoxyribonucleic acid
GC content	the content of guanine and cytosine base pairs in a DNA sequence. Normally reported as a percentage of total base pairs
GC%	percentage of guanine and cytosine base pairs in a DNA sequence
indel	insertion/deletion
infA	translation initiation factor IF-1
IR	inverted repeat
IR_A	inverted repeat copy A
IR_B	inverted repeat copy B
ITS1	internal transcribed spacer 1
ITS2	internal transcribed spacer 2
kb	kilobase
LSC	large single copy
matK	maturase K gene
MGS	Massey Genome Service
ML	maximum likelihood
mm	millimetre
mya	million years ago
NADH	nicotinamide adenine dinucleotide
ndh	nicotinamide adenine dinucleotide (NADH) dehydrogenase subunit genes. Individual subunit genes are denoted A-K (i.e., *ndhA-ndhK*)
nrITS	nuclear ribosomal internal transcribed spacer
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
pet	cytochrome b/f complex subunit encoding gene
psa	photosystem I subunit genes. Individual subunit genes are denoted A-E, I-M (i.e., *psaA-psaM*).
psb	photosystem II subunit encoding gene. Individual subunit genes are denoted A-F, H-N (i.e., *psbA-psbN*).
rbcL	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene
recA	chloroplast-targeted homologue of the nuclear-encoded DNA repair protein gene
redox	reduction-oxidation reactions
RNA	ribonucleic acid
rpl	ribosomal protein large subunit genes. Individual subunit genes are denoted 2, 14, 16, 20, 22, 23, 32, 33, 36 (i.e., *rpl2-rpl36*)
rpo	plastid-encoded RNA polymerase subunit genes. Individual subunit genes are denoted A-C2 (i.e., *rpoA-rpoC2*).
rps	ribosomal protein small subunit genes. Individual subunit genes are denoted

	2-4, 7, 8, 11, 12, 14-16, 18 and 19 (i.e., *rps2-rps19*)
rrn	ribosomal RNA genes
rRNA	ribosomal RNA
SNP	single nucleotide polymorphism
spp.	species
SSC	small single copy
trn	transfer RNA gene
tRNA	transfer RNA
trnE	transfer RNA gene for glutamic acid
trnK	transfer RNA gene for lysine
trnL	transfer RNA gene for leucine
trnT	transfer RNA gene for threonine
ycf	hypothetical chloroplast RF gene. Individual *ycf* genes are denoted 1-4 and 15 (i.e., *ycf1- ycf4* and *ycf15*).

Chapter 1

Introduction

1.1 Plastids and their genomes

Merechowsky (1905; translated by Martin & Kowallik, 1999) described chloroplasts as a "plethora of little workers, green slaves" (pg. 292 Martin & Kowallik, 1999) whose role was to nourish the plant cells they reside in by producing sugars from sunlight. Chloroplasts use photosynthesis to convert light energy to chemical energy. This physiological process is critical to the survival of green plants and, ultimately, one of the most important biochemical processes on Earth.

The endosymbiont theory is now widely accepted as explaining the origin of plant plastids including chloroplasts, amyloplasts and others (Margulis, 1970; Martin & Kowallik, 1999). This hypothesis suggests that plastids are derived from a photosynthetic, cyanobacterial-like prokaryote that was engulfed by a non-photosynthetic, eukaryotic cell. Rather than being digested, these prokaryotes became endosymbionts. Over evolutionary time, they have become tightly integrated into the host cells. Now, as organelles, they confer the ability to photosynthesise (Margulis, 1970; Martin & Kowallik, 1999). Several lines of evidence support the endosymbiont theory. These include, similarities between the structure and physiology of plastids and free-living cyanobacteria (Fairman et al., 2011); the double membrane structure of plastids (Whatley & Whatley, 1981); that plastids divide by binary fission and independently of the plant cell (Margolin, 2005; Miyagishima, 2011; Pyke, 1999); that the ribosomes of plastids are more like bacterial than eukaryotic ribosomes (Ellis, 1970; Manuell et al., 2007), and that, like cyanobacteria, plastids have their own circular genomes (Timmis et al., 2004).

Extant plastid genomes, or plastomes, are distinct from the nuclear and mitochondrial genomes of plants and differ from those of modern cyanobacteria. The three plant genomes each occupy different cellular compartments and differ in their evolutionary origins. Although plastomes and cyanobacterial genomes have shared ancestry, the environments they occupy imply different functional contraints and, therefore, evolutionary trends. Consistent with the internal environment of a plant

1

cell being relatively stable, the evolution of the plastome has been characterised by loss of genes that may have been necessary for the survival of the cyanobacterial ancestor in less stable, extracellular environments. As a result, the plastomes of extant land plants are far smaller and encode far fewer genes than the genomes of their free-living cyanobacterial relatives (Dyall et al., 2004; Palmer, 1985). Specifically, the genomes of modern cyanobacteria encode several thousand genes whereas the plastomes of land plants encode 60-80, and those of eukaryotic algae (e.g., red algae) up to 200 genes. It is therefore likely that the plastomes of contemporary land plants and eukaryotic algae encode less than 10% of their original gene complement (Dyall et al., 2004; Martin & Herrmann, 1998; Martin et al., 2002).

Although discarded from the plastome, not all these genes have been entirely lost. Instead, transfer of genetic material to the nuclear and mitochondrial genomes has resulted in the retention of much of this functionality (Kleine et al., 2009). For example, in *Arabidopsis thaliana* (L.) Heynh (Brassicaceae) up to 18% of the protein-coding nuclear genes may have cyanobacterial origins. The products of at least some of these are subsequently exported to the plastids (Daniell et al., 2016; Martin et al., 2002). Indeed, potentially many of the protein products required for chloroplast function are nuclear-encoded. Current estimates suggest up to 5000 proteins may be imported by plastids; presumably many are encoded by genes that were transferred to the nucleus during plastome reduction (Peltier et al., 2000). In plant cells, integration of the chloroplast, mitochondrial, and nuclear genomes is central to co-ordination of core cellular processes such as photosynthesis and respiration (Kleine et al., 2009; Levin, 2003). For example, in some interspecific hybrids reduced fitness has been attributed to disruption of the integration between parental organelle and nuclear genomes (Burton et al., 2013).

1.2 The plastomes of extant photosynthetic plants

The plastomes of almost all extant photosynthetic land plant lineages are highly similar with respect to overall size, structure, and gene content (Daniell et al., 2016; Palmer, 1985; Xu et al., 2015). In photosynthetic land plants, plastid genomes are commonly 107-218 kb in size with a quadripartite structure consisting of a large (LSC) and small (SSC) single copy region separated by a pair of inverted repeats (IR) (Figure 1.1) (Daniell et al., 2016; Palmer, 1985). In total 120-130 genes are

encoded by the plastome; this includes a number that lie within the IR and therefore have two copies per genome (e.g., the ribosomal subunit genes). These genes have roles in photosynthesis (e.g., photosystem I and II genes), transcription and translation (e.g., the ribosomal protein genes) and housekeeping roles (e.g., a gene encoding the proteolytic subunit of the ATP-dependent Clp protease and one encoding a maturase involved in group II intron splicing) (Daniell et al., 2016; Graham et al., 2017).

Many of the features typical of extant photosynthetic land plant plastomes (e.g., gene content and quadripartite structure) are thought to have been present in the Streptophyte lineage from which land plants evolved (de Cambiaire et al., 2006) This implies that the land plant plastome must have remained relatively stable since these groups diverged (Turmel et al., 2005, 2006, 2007). Also consistent with stability over time, in most extant land plant lineages the IR, LSC, and SSC regions of the plastome typically contain the same core gene sets (Mower & Vickrey, 2018). Nevertheless, minor variations are common and primarily involve differences in overall size, the relative contributions of the LSC, SSC, and IRs to plastome size, and gene content (Mower & Vickrey, 2018). For example, the eleven plastid NADH dehydrogenase (*ndh*) genes are frequently absent, particularly from photosynthetic Orchidaceae plastomes (Kim & Chase, 2017). Major departures from the common land plant plastome structure have also been found in a few lineages but are less common. For example, only three photosynthetic angiosperm lineages have been found to lack the IR; the genus *Erodium* Aiton (Geraniaceae; Blazier et al., 2016), a monophyletic subgroup of legumes (Fabaceae; Palmer & Thompson, 1981) and the saguaro cactus, *Carniegia gigantea* (Engelm.) Britton & Rose (Cactaceae; Sanderson et al., 2015).

Several observations have been linked to lower substitution rates (Wolfe et al., 1987; Xu et al., 2015) and structural stability (Daniell et al., 2016; Palmer, 1985; Xu et al., 2015) in plastomes of photosynthetic land plants. One suggestion is that the characteristic quadripartite structure influences plastome stability, and specifically, that the length and nature of the IRs could impact homologous recombination (Palmer & Thompson, 1981; Strauss et al., 1988). That substitution rates differ for genes located within IRs (c.f., those in the single copy regions) has been suggested as evidence of the IRs role in plastome stability (Perry & Wolfe, 2002; Wolfe et al., 1987). Alternatively, the conservative nature of the plastome could be a consequence
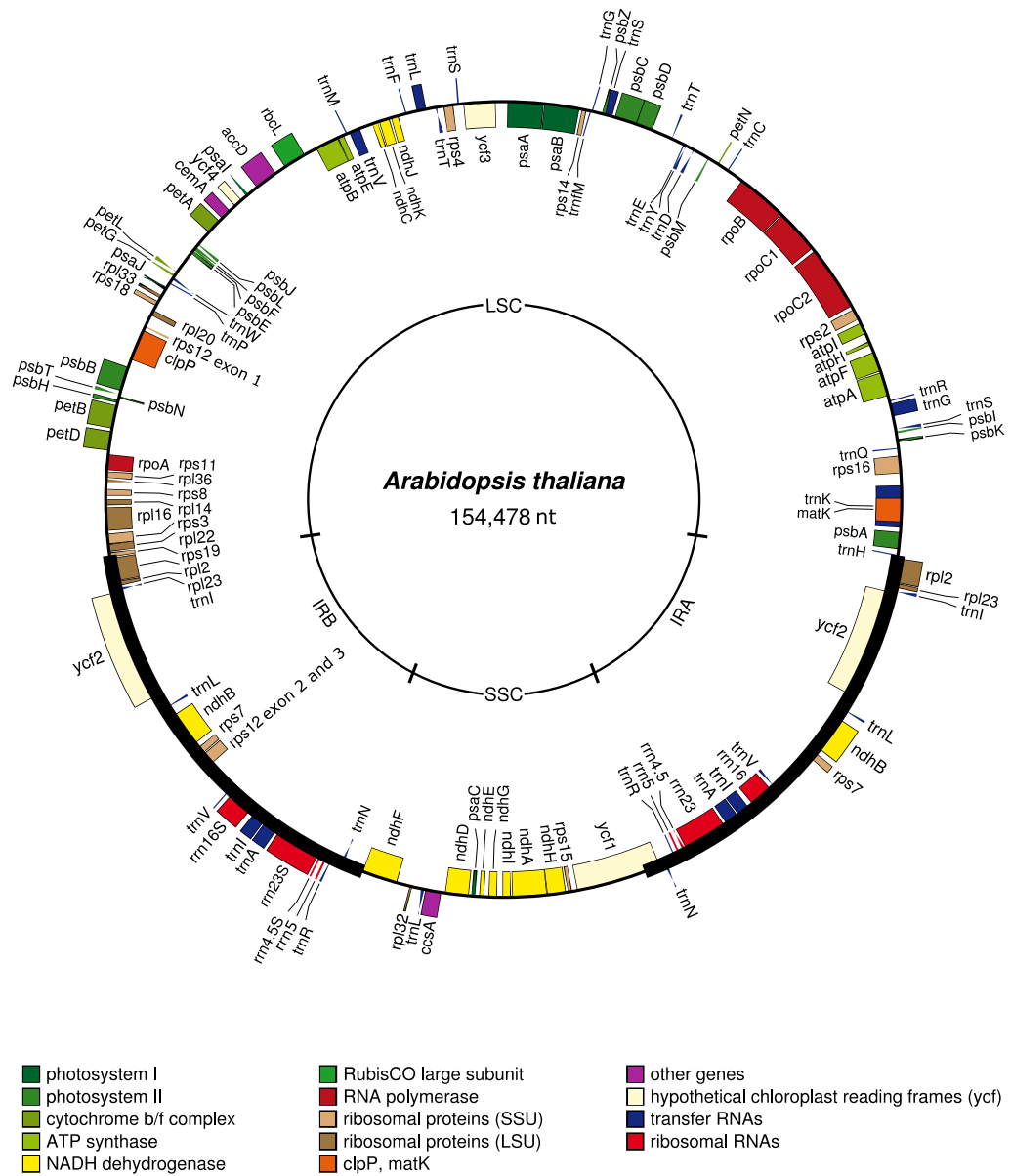
Figure 1.1 Plastome of the photosynthetic angiosperm *Arabidopsis thaliana* (L.) Heynh (Brassicaceae). The gene content, gene order, and four-part plastome structure are typical of photosynthetic land plants. The inner circle indicates the extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Genes on the interior of the outer circle are transcribed in the forward direction; those on the exterior are transcribed in the reverse direction. Plastome (Sato et al., 1999) drawn using OrganellarGenomeDraw (Lohse et al., 2013).

of specific mechanisms. For example, since no plastome-encoded proteins appear to function in DNA repair it is assumed that repair of the chloroplast genome requires nuclear-encoded proteins to be imported (Maréchal & Brisson, 2010; Rowan et al., 2010). Consistent with this assumption, disruption of the chloroplast-targeted

homologue of the nuclear-encoded DNA repair protein recombinase A (*recA*) gene is associated with aberrant plastomes in *A. thaliana* (Rowan et al., 2010). Although ultimately the mechanism remains uncertain, increased genome instability is coupled with loss of the quadripartite structure in some legumes (Palmer & Thompson, 1981), conifers (Strauss et al., 1988), and green algae (Bélanger et al., 2006; Turmel et al., 2005).

Other hypotheses suggest that stability of the plastid genome reflects a requirement that certain genes are retained (Allen, 2003). These genes may be retained because their transfer to another genome is simply too difficult (i.e., they have become "trapped"). This might occur if the protein product encoded by the gene was particularly hydrophobic and therefore difficult to transport from the cytoplasm into the plastid. This suggestion appears to explain some retained genes but not others (Daley & Whelan, 2005; Kleine et al., 2009). Alternatively, retention may reflect functionality (i.e., genes are selectively retained) (Daley & Whelan, 2005; Graham et al., 2017). One example is the co-location for redox regulation (CoRR) hypothesis. This suggests that the ability to rapidly adjust the electron transport chain depends on the genes involved remaining within the same membrane-bound compartment. If so, this would reduce the likelihood of genes associated with the electron transport chain being transferred from the plastome to another genome (Allen, 1993, 2003; Pfannschmidt, 2003).

These phenomena – overall genome structure, gene function constraints, and the difficulty of gene transfer – are not mutually exclusive and may act collectively to explain the apparent stability of extant photosynthetic plant plastomes.

1.3 Heterotrophic plants and their plastomes

Not all land plants are photosynthetic. Some are heterotrophic, meaning they cannot produce their own carbohydrates and instead obtain organic compounds via a host (Merckx, 2012; Nickrent, 2002). In some cases, these plants are strict heterotrophs; these have lost the ability to produce chlorophyll and therefore obtain all necessary organic compounds from their host. In other cases, organic compounds come from a combination of the plant's own photosynthesis and from their host (Julou et al., 2005; Merckx, 2012; Nickrent, 2002).

Two broad types of heterotrophic plants are recognised. The parasitic forms meet some or all of their requirements for nutrition or water directly from a plant host (Nickrent, 2002; Press et al., 1999). There are an estimated 4,100 plant species with a parasitic lifestyle, of which 390 are thought to be strict heterotrophs (or holoparasites) (Merckx, 2012). Well known examples of parasitic plants include the corpse flowers (*Rafflesia* R. Br., Rafflesiaceae), the dodders (*Cuscuta* L., Convolvulaceae), and the mistletoes (Viscaceae and Loranthaceae) (Nickrent, 2002). The second form of heterotrophy in plants is mycoheterotrophy. Although more common this form is more poorly studied, in part because this lifestyle has only been recognised for about 25 years (Leake, 1994; Merckx, 2012). Currently, around 515 strictly mycoheterotrophic (or holomycoheterotrophic) angiosperms are known. Belonging to 10 plant families, these rely on a fungal associate to provide all necessary organic compounds and typically lack leaves and chlorophyll.

Estimates suggest that as many as 20,000 plant species are partial mycoheterotrophs. These plants retain their chlorophyll and augment photosynthesis-derived organic compounds with those from a fungal associate for at least part of their lifecycle (Leake, 1994; Merckx, 2012). The mycoheterotrophic lifestyle appears to have evolved on multiple occasions, likely from situations where the relationship was initially symbiotic. These so-called mycorrhizal associations, involving a mutually beneficial relationship between a plant and soil fungus, occur in ~82% of land plants. Typically, the plant partner provides photosynthetically-derived organic compounds to the fungus in exchange for nutrients (Brundrett, 2002).

Despite the loss of chlorophyll in some heterotrophic plants, all those studied to date retain chloroplasts. However, the shift to heterotrophy is associated with plastome changes. Compared to photosynthetic plants, the plastomes of heterotrophic plants are typically smaller, encode fewer genes, and are structurally rearranged (Graham et al., 2017) (Figure 1.2). The extent to which the plastome is impacted differs markedly among species. At one end of the spectrum, the plastomes of some heterotrophic plants differ little from those of autotrophic plants for size and gene content. For example, the chloroplast genome of the mycoheterotroph *Cymbidium macrorhizon* Lindl. (Orchidaceae) is 149.85 kb long. This is within the typical range for an autotrophic plant (Kim et al., 2018). Gene content in *C. macrorhizon* is also similar to that of an autotrophic plant plastome. In this case, genes encoding several of the *ndh* genes have been lost and that encoding a hypothetical chloroplast protein
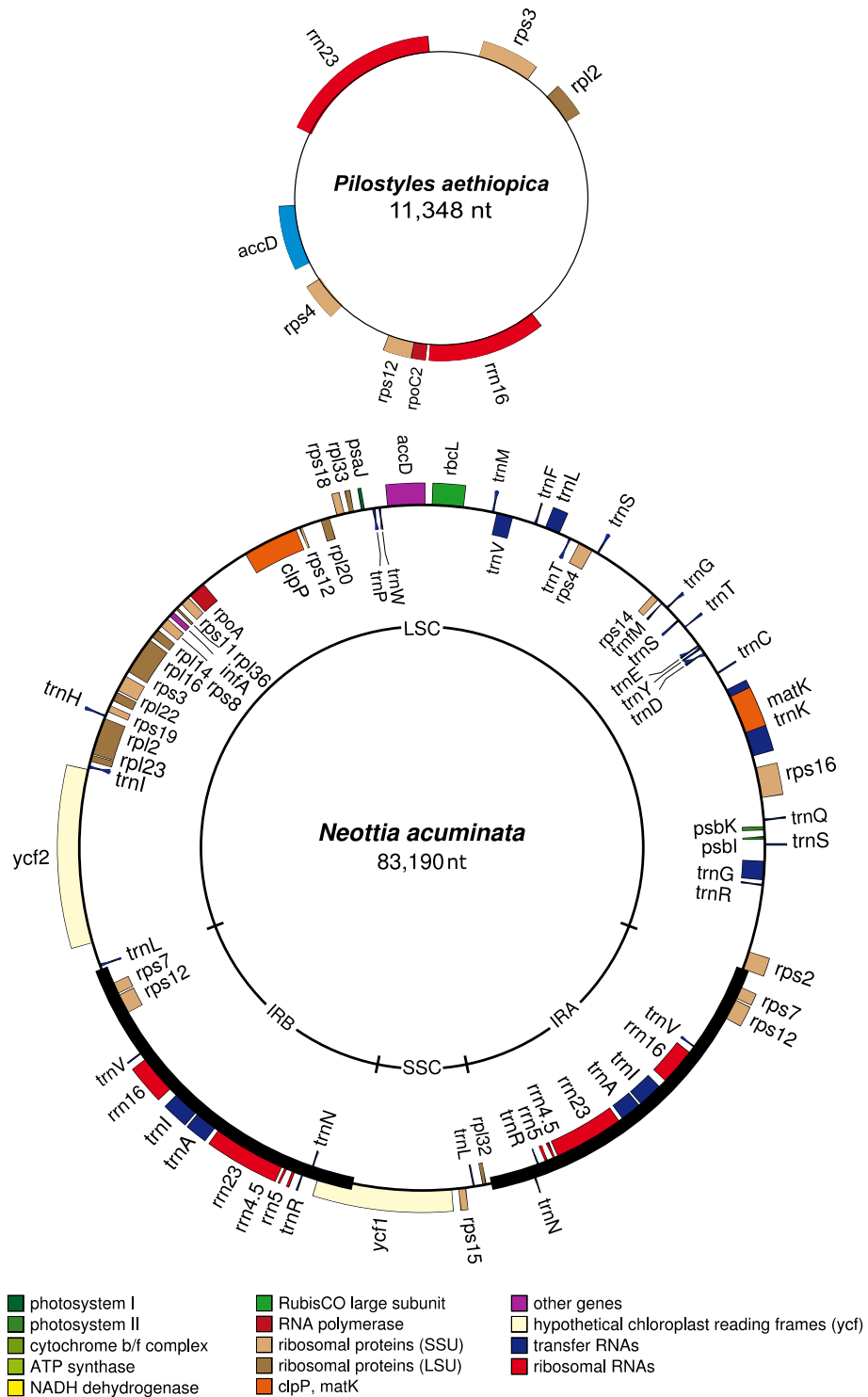
Figure 1.2 Plastomes of holoparasitic *Pilostyles aethiopica* Welw. (Apodanthaceae) and holomycoheterotrophic *Neottia acuminata* Schltr. (Orchidaceae). For both plastomes, genes are represented as coloured blocks; genes on the interior of a circle are transcribed in the forward direction, those on the exterior in the reverse direction. For *N. acuminata* the inner circle indicates locations of the inverted repeat (IR_A and IR_B), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle represents the IR regions. Plastomes from Bellot and Renner (2016); and Feng et al. (2016); drawn using OrganellarGenomeDraw (Lohse et al., 2013).

(*ycf15*) has been pseudogenised. At the other end of the spectrum, plastomes can be highly reduced. The smallest known plant plastome is that of the parastic plant *Pilostyles aethiopica* Welw. (Apodanthaceae). This plastome is only 11.3 kb in length and encodes five or six genes. For such smaller plastomes, structural rearrangements are almost ubiquitous (e.g., Feng et al., 2016; Schelkunov et al., 2015).

It is unclear whether the shift to heterotrophy, and more specifically mycoheterotrophy, precedes or follows plastome change. One suggestion is such a trophic shift may follow a change in fungal partner (Ogura-Tsujita et al., 2012). Rather than replacing one symbiotic relationship with another, the plant exploits a new fungal partner for water, nutrients, and as a carbon source. In this case it would no longer be necessary for the plant to produce its own food and any selection acting to maintain the plastome would be reduced. Presumably at that point changes to the plastome (e.g., substitutions and structural rearrangements) would accumulate rapidly. Consistent with this possibility the fungal associates of some mycoheterotrophs differ from those of their photosynthetic relatives (Merckx, 2012; Merckx et al., 2009; Ogura-Tsujita et al., 2012). For example, commonly the mycobionts of the mycoheterotrophic *C. macrorhizon* are members of the Sebacinales whereas autotrophic species of *Cymbidium* Swartz are associated with members of the Tulasnellaceae. Partially mycoheterotrophic *Cymbidium* species have associations with various fungi, including members of both the Tulasnellaceae and Sebacinales (Ogura-Tsujita et al., 2012).

This example also provides insight into the relative timing of changes. Like many mycoheterotrophs, the vegetative body of *C. macrorhizon* is a subterranean rhizome with highly reduced scale leaves; this growth form would appear to have little need for photosynthesis. Yet the aboveground portion of *C. macrorhizon*, the ephemeral flowering stem, is pale green and has been shown to be photosynthetic (Suetsugu et al., 2018). Consistent with this latter observation, the plastome of *C. macrorhizon* is highly similar to those of photosynthetic *Cymbidium* species in terms of size and gene content (Kim et al., 2018). At least for this species, the morphological, physiological and life history changes associated with a shift to mycoheterotrophy have occurred prior to plastome degradation. This scenario is also consistent with evidence suggesting that retention of the photosynthetic progenitor's

morphology (e.g., leaves) may be selectively disadvantagous in non-photosynthetic decendents (e.g., Julou et al., 2005).

An alternative is that loss of photosynthesis precedes the transition to mycoheterotrophy. Since both the chloroplast and photosynthesis rely on protein products encoded in other genomic compartments, the loss of function may reflect changes to the chloroplast, mitochondrial or nuclear genomes. While a loss of photosynthesis would appear to be unsurvivable for a green plant, survival might be possible if an existing mutualistic, mycorrhizal association could be subverted to also act as a carbon source. Observations on *Cephalanthera longifolia* (L.) Fritsch (Orchidaceae) are consistent with this possibility. Specifically, in this species non-photosynthetic albino individuals occasionally arise and may persist for several years alongside green, presumably photosynthetic individuals (Abadie et al., 2006). Both albino and photosynthetic individuals have the same mycobiont, suggesting that the loss of photosynthesis may change the nature of an existing plant-fungal relationship (Julou et al., 2005). What remains uncertain is whether such cases represent a stepping-stone toward mycoheterotrophy.

Regardless of causality, a shift to heterotrophy is typically associated with a suite of changes to the plastome. These changes are often characterised as resulting in plastome "degradation" (Barrett & Davis, 2012; Barrett et al., 2014; Graham et al., 2017). In heterotrophic plants the chloroplast no longer performs a key function and, presumably, the selective forces maintaining the plastome are weakened. That almost all the heterotrophic plants studied to date have been found to retain a plastomes in some form suggests that these genomes may persist long after loss of photosynthesis. Conversely, the failure to recover a complete plastome sequence from the holoparasite *Rafflesia lagascae* R. Br. (Rafflesiaceae) has led to the suggestion that given sufficient time plastomes may be lost entirely (Molina et al., 2014).

A key question is why plastomes persist in heterotrophic plants if most plastid-encoded genes are directly or indirectly involved in photosynthesis. One possibility is that although the plastomes of non-photosynthetic plants are degrading, there has simply not been enough time for their complete loss in most cases. While certainly a possibility, this view overlooks the fact that plastids have roles in processes other than photosynthesis. For example, in gravity perception and storage (Inaba & Ito-Inaba, 2010; Moctezuma & Feldman, 1999). A few plastome-encoded gene products function in biochemical processes aside from photosynthesis (e.g.,

haeme biosynthesis; Smith and Lee, 2014) as well as in general maintenance of the chloroplast (e.g., the ATP-dependent Clp protease; Graham et al., 2017). It has also been suggested that genes involved in general maintenance – the so-called housekeeping genes – would be difficult to transfer to the nucleus (Delannoy et al., 2011; Graham et al., 2017). These observations are consistent with the idea that the need to retain at least a few plastid genes would drive selection for the maintenance of the plastome long after the loss of photosynthesis.

Rather than heading towards complete loss, it has been suggested that the plastomes of heterotrophic plants are converging on a minimal gene set (Wolfe et al., 1992). That is, a gene set that cannot be lost from the plastome without risking plant survival. Several genes have been suggested as potential members of this minimal set, primarily because they had been retained in sequenced heterotrophic plant plastomes. These included genes for the proteolytic subunit of the ATP-dependent Clp protease (*clpP*), acetyl-coenzyme A carboxylase (*accD*), and the transfer RNA for glutamic acid (*trnE*) (Barrett & Davis, 2012; Barrett et al., 2014; Graham et al., 2017). However, with the increasingly rapid accumulation of plastome sequences all these genes are now known to be absent from the plastome of at least one species. For example, the *trnE* gene is absent in two holoparasitic *Pilostyles* Guill. species (Bellot & Renner, 2016) and the *clpP* gene is absent in the mycoheterotroph, *Thismia tentaculata* K. Larsen & Aver (Burmanniaceae) (Lim et al., 2016). Although it now seems unlikely that all heterotrophic plant plastomes are converging toward a single minimal gene set, the idea of minimal sets may still prove useful. One possibility is that lineages are converging towards minimal gene sets that differ depending upon the interactions between genomic compartments (Schelkunov et al., 2015).

1.4 A model of plastome reduction in heterotrophic plants

The pathway of plastome degradation in heterotrophic and, particularly, mycoheterotrophic plants is a topic of much interest (Merckx, 2012). Genome comparisons at the intergeneric and interfamilial levels suggest that heterotrophy has arisen independently on multiple occasions and that subsequent evolution of the plastome differs in terms of gene losses and genome rearrangements between lineages (Barrett & Davis, 2012; Delannoy et al., 2011; Feng et al., 2016; Wicke et al., 2013). As a result, there is greater interspecific and intergeneric variation among

the plastomes of heterotrophic than autotrophic plants. Despite the apparent differences between taxa (Barrett & Davis, 2012; Delannoy et al., 2011; Feng et al., 2016; Wicke et al., 2013), it has been suggested that there is a broad, functionally determined pattern to plastome degradation following a shift to mycoheterotrophy (Barrett & Davis, 2012; Barrett et al., 2014; Graham et al., 2017).

Barrett and Davis (2012) have formalised the general sequence of degradation as a stepwise model that links the pattern of gene loss to gene function. Their model has five distinct stages. The first stage involves loss of the *ndh* genes. The NADH dehydrogenase protein complex is typically encoded by 11 genes and has roles in maximising photosynthetic efficiency as well as in photoprotection (Endo et al., 1999; Leonid et al., 1998). These genes are also absent from the plastomes of some photosynthetic species; either they are not vital for photosynthesis or their functionality is easily acquired by another genomic compartment (Chang et al., 2006; McCoy et al., 2008; Wakasugi et al., 1994). Consistent with the observation that most mycoheterotrophic plants occupy forest understorey habitats where the role of *ndh* complex may be of limited value, some or all of the *ndh* genes have been lost in most mycoheterotrophic plants (Barrett et al., 2014).

In the second stage of the model, the reduced importance of photosynthesis is assumed to relax purifying selection on photosynthesis-related genes thereby allowing mutations to accumulate and, ultimately, leading to their loss (Barrett & Davis, 2012). Genes lost at this stage include those encoding subunits of the light-harvesting photosystems I and II (*psa* and *psb* complex genes, respectively), those encoding the cytochrome b6f complex (*pet* genes) involved in photosynthetic electron transport, and the large subunit of the carbon-fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) (Graham et al., 2017).

The third stage proposes loss of the plastid-encoded RNA polymerase (*rpo*) genes. This polymerase has a role in transcribing operons that encode photosynthesis-related products (Daniell et al., 2016). Loss of photosynthesis genes in the previous stage is suggested to relax selection on the *rpo* genes and, ultimately, lead to their loss. In mycoheterotrophic plants where all or most of the photosynthesis genes have been lost, the *rpo* genes are also typically absent (Barrett & Davis, 2012; Daniell et al., 2016).

The fourth and fifth stages of the model involve loss of genes with roles beyond photosynthesis and are therefore likely to be retained for longer. The ATP

synthase subunit (*atp*) genes are lost in the fourth stage. Although this complex plays a role in photosynthesis, its retention may reflect functions in processes associated with other plastid types (Barrett & Davis, 2012). Finally, the most persistent are the housekeeping genes. These are involved in maintenance of the plastid itself, including protein synthesis, RNA processing, and protein degradation/recycling (Barrett & Davis, 2012; Graham et al., 2017). These genes include the transfer RNA (*trn*) genes and maturase K (*matK*).

The Barrett and Davis model (Barrett & Davis, 2012) describes observations on the available plastome sequences for mycoheterotrophic plants. It is tempting to assume that this model provides a general explanation of plastome degradation in these groups. That said there are several reasons to remain cautious. First, the number of complete plastome sequences for heterotrophic plants remains small and taxonomic coverage is limited (Logacheva et al., 2014). Around 515 holomycoheterotrophic angiosperm species are currently known but the plastomes of only 20, representing five of the 10 holomycoheterotroph-containing families, have been sequenced. Sampling is also biased. Although the majority of holomycoheterotrophs are monocots – seven of the 10 recognised holomycotroph containing plant families are monocots and one of those, Orchidaceae, contains ~235 species – just 3% have complete plastome sequences, compared to 13% of the eudicot holomycoheterotrophs. The Ericaceae is currently the best sampled holomycoheterotrophic containing plant family with complete plastome sequences available for five of its ~18 (28%) holomycoheterotrophic members (Logacheva et al., 2016). In contrast only 11 of the ~235 (5%) holomycoheterotrophic Orchidaceae have complete plastome sequences available.

Second, the Barrett and Davis model (Barrett & Davis, 2012) focuses exclusively on functionally-led loss; it implies that size reductions and structural rearrangements in heterotrophic plant plastomes are an outcome of progressive gene loss. One alternative is that structural changes, viewed as an outcome by Barrett and Davis (2012), have initiated the process of genome reduction, if not the shift to heterotrophy itself. It is well established that plastome structure is much more variable for heterotrophic than autotrophic species (Blazier et al., 2016; Lavin et al., 1990; Wakasugi et al., 1994). In autotrophic plants structural changes, especially large ones, are likely to have deleterious effects and, therefore, unlikely to be tolerated (Inaba & Ito-Inaba, 2010). However, assuming an appropriate fungal

association is already established, such effects could conceivably be tolerated in mycoheterotrophs.

In non-photosynthetic plant plastomes characterised to date, large-scale deviations from the typical land plant plastome architecture are not uncommon. These include, large expansions and contractions of the IR, loss of the IR, and loss of some or all the SSC region. For example, although the plastomes of *Rhizanthella gardneri* R. S. Rogers (Orchidaceae), *Epipogium roseum* (D. Don) Lindl. (Orchidaceae) and *Petrosavia stellaris* Becc. (Petrosaviaceae) all retain the typical quadripartite structure, each has been altered in a unique way (Delannoy et al., 2011; Logacheva et al., 2011; Schelkunov et al., 2015). In both *R. gardneri* and *E. roseum* the IR has been reduced; in *R. gardneri* the ribosomal RNA (*rrn*) genes fall within the SSC region instead of the IR (Delannoy et al., 2011), whereas in *E. roseum* only the transfer RNA gene for leucine (*trnL*) remains within the IR (Schelkunov et al., 2015). In *P. stellaris*, gene order has been altered (Logacheva et al., 2014). In some heterotrophic plant species the quadripartite structure has been lost. For example, in contrast to *E. roseum*, the IRs have expanded to encompass the entire SSC region in *E. aphyllum* Sw. (Schelkunov et al., 2015). One copy of the IR has been entirely lost from *Monotropa hypopitys* L. (Ericaceae) (Ravin et al., 2016). It has also been suggested that smaller structural elements, such as short dispersed repeats, could act to destabilise the plastome (Strauss et al., 1988).

The Barrett and Davis model (Barrett & Davis, 2012) also does not consider possible interactions among these different types of change. Given the limits of the available data, it is difficult to determine whether changes attributed to any one of these processes – functional loss, large-scale structural change, or instability as the result of small structural repeats – initiates or is a consequence of the other two. Indeed, if outcomes are viewed in terms of relaxed selection, then changes initiated by any one of these three processes could result in changes caused by either of the other two. Moreover, it is possible that the sequence of events differs between species. In some, functional loss may be the primary driver and in others, it may be structural change.

Given currently available sampling, only wide comparisons are possible even within groups such as the Ericaceae and Orchidaceae. Such comparisons are not ideal for evaluating the sequence of plastome change for any given lineage, and therefore support for the Barrett and Davis (2012) model remains limited. A fuller evaluation

of the Barrett and Davis model will require the plastomes of many more heterotrophic plants to be characterised.

1.5 New Zealand Orchidaceae as a study group

Currently, too few full plastome sequences are available for mycoheterotrophic plant species to draw strong conclusions about the relative importance of functional loss and structural change in the evolution of these genomes. Increasing the sample of plastome sequences for mycoheterotrophic plants has the potential to provide new evolutionary insights.

The Orchidaceae are a particularly important target for additional plastome sequencing as this large, diverse family contains approximately half of the currently known mycoheterotrophic plant species (Merckx, 2012). To date, plastomes from just 18 taxa of predominantly European and North American mycoheterotrophic orchids have been characterised (Barrett & Davis, 2012; Barrett et al., 2014; Delannoy et al., 2011; Feng et al., 2016; Kim et al., 2018; Logacheva et al., 2011; Schelkunov et al., 2015; Yuan et al., 2018). Based on the distribution of holomycoheterotrophy across Orchidaceae, this lifestyle is estimated to have arisen at least 25 times within the family. This is likely to be an underestimate, since holomycoheterotrophy is now thought to have arisen on multiple occasions in some genera (Barrett et al., 2014; Feng et al., 2016; Merckx, 2012). Moreover, the importance of holomycoheterotrophy differs between groups. In some cases all members of a genus are non-photosynthetic (e.g., *Aphyllorchis* Blume, *Gastrodia* R. Br.; Merckx, 2012) while others include both photosynthetic and non-photosynthetic members (Feng et al., 2016; Merckx, 2012). Among the latter are examples of multiple, independent origins within a singe genus (e.g., *Corallorhiza* Gagnebin; Barrett and Davis, 2012; Barrett et al., 2014), genera containing distinct subclades of photosynthetic and non-photosynthetic species (e.g., *Neottia* Guett.; Feng et al., 2016), and genera with just a single holomycoheterotroph (e.g., *Corybas* Salisb., *Platanthera* (L.) Rich.; Lyon, 2015; Merckx, 2012). The available plastome sequences simply do not fully represent this diversity.

In expanding the collection of plastome sequences there is obviously a need to include a broader taxonomic and geographic range of mycoheterotrophs; these may provide examples of additional evolutionary patterns or highlight patterns that

are common. It will also be important to expand our sampling of the mycoheterotroph-containing lineages. Although wide comparisons are important in terms of overall patterns, it may only be possible to understand key evolutionary processes using comparisons among more closely related species (whether photosynthetic or not).

New Zealand has a rich orchid flora, comprising 34 genera and over 100 species (Breitwieser et al., 2012; Dawson et al., 2007). Of these, just *Corybas*, *Danhatchia* (Hatch) Garay & Christenson, and *Gastrodia* contain non-photosynthetic, mycoheterotrophic representatives (Lehnebach et al., 2016b; Merckx, 2012; Moore & Edgar, 1970; St. George, 1999) (Figure 1.3). These genera differ in terms of species diversity and biogeographical origins. *Corybas* is a genus of ~135 species widely distributed across Asia, Australia, New Zealand and the Pacific Islands (Lyon, 2015; Merckx, 2012). There are 21 species currently recognised in New Zealand, including the sole non-photosynthetic, mycoheterotrophic member of the genus, *Corybas cryptanthus* Hatch (Hatch, 1956; Lyon, 2015; Merckx, 2012; Moore & Edgar, 1970). In contrast, *Gastrodia* contains ~70 non-photosynthetic, mycoheterotrophic species. This genus is relatively widespread across Asia, Africa and the Pacific Islands with five species recorded from New Zealand (Lehnebach et al., 2016a; Merckx, 2012; Moore & Edgar, 1970). Finally, until recently *Danhatchia* was considered monotypic, endemic to New Zealand (Banks, 2012; Merckx, 2012; Moore & Edgar, 1970). However, the genus has now also been reported from Australia (Banks, 2012; Jones & Clements, 2018; Steenbeeke, 2012).

Corybas, *Danhatchia*, and *Gastrodia* differ in terms of their total diversity and relative proportions of mycoheterotrophs as well as how closely related these are to photosynthetic representatives. In comparative analyses these differences have the potential to lead to new insights into the evolution of mycoheterotrophic plastomes. For example, investigating intraspecific plastome variation in genera where mycoheterotrophy is likely to have originated at different times may provide insights into rates and patterns of change. Similarly, comparisons with photosynthetic relatives in groups where these are closely or more distantly related could provide insights into the drivers of change. Such studies may also provide data that would allow the Barrett and Davis (2012) model of plastome degradation to be tested. Beyond questions about plastome evolution these groups may offer a framework for
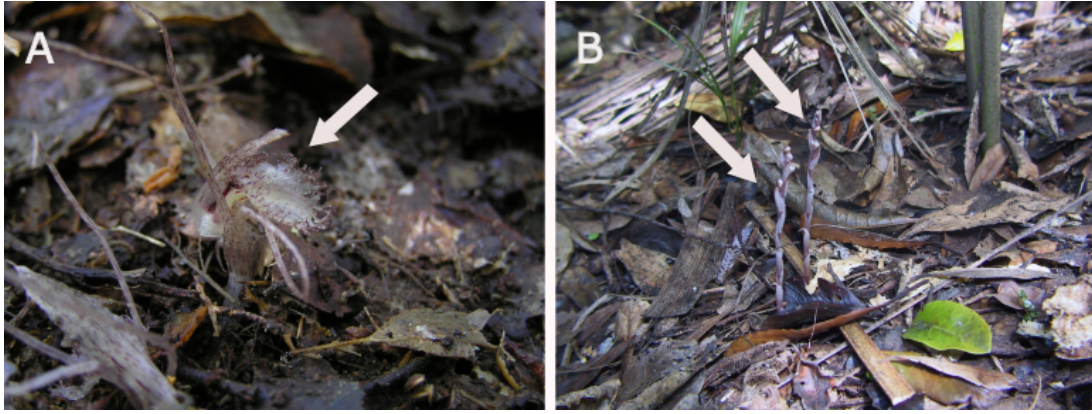
Figure 1.3 New Zealand mycoheterotrophic orchids *in situ*. A. *Corybas cryptanthus* Hatch. B. *Danhatchia australis* (Hatch) Garay & Christenson, two specimens. Photographs taken by Michael Moorhouse.

investigating the evolution of mycoheterotrophy. For example, that the transition to mycoheterotrophy involves a shift in the fungal associate.

1.6 Focus of the investigation

This thesis describes the characterisation and analyses of plastomes from two holomycoheterotrophic New Zealand orchids; *C. cryptanthus* and *Danhatchia australis* (Hatch) Garay & Christenson. The plastome of a recently discovered albino *Corybas* specimen is also sequenced. Using these data, a pair of hypotheses related to the Barrett and Davis (2012) model of the plastome degradation in mycoheterotrophs are considered.

First, I hypothesise that the plastomes of these three putatively non-photosynthetic taxa will be reduced in terms of size and gene content relative to those of photosynthetic relatives. However, I predict that the extent of plastome reduction will differ between them. Specifically, I predict that the plastome of the apparently taxonomically isolated *D. australis* will be more reduced than that of *C. cryptanthus*, for which close photosynthetic relatives are recognised. Given that the albino *Corybas* retains a typically photosynthetic morphology I predict that this plastome will be more intact than that of either *C. cryptanthus* or *D. australis*.

Second, I hypothesise that patterns of gene loss from the plastomes of
C. cryptanthus, *D. australis*, and the albino *Corybas* will be consistent with the
model proposed by Barrett and Davis (2012). That is, there will be stepwise
disruption of functional gene classes. Disruption of the *ndh* genes will be followed
by that of the so-called photosynthetic genes, with the *rpo*, *atp*, and housekeeping
gene classes being disrupted at progressively later stages. Disruption of genes
belonging to any one of these functional classes should begin only after the
functional or physical loss of most or all the genes from the class immediately prior.
Related to this, I further predict that the plastomes of photosynthetic relatives of
C. cryptanthus, *D. australis*, and the albino *Corybas* will exhibit patterns of sequence
evolution that are consistent with these groups being predisposed to the loss of
photosynthesis and plastome degradation. For example, that there will be elevated
levels of sequence variability in the *ndh* genes of photosynthetic *Corybas* consistent
with reduced selection prior to initiation of gene loss or that the plastomes of
photosynthetic relatives will be structurally modified in ways that could potentially
destabilise these genomes.

The next two chapters describe analyses of the plastomes of *C. cryptanthus*
and *D. australis*. Chapter 2 describes plastome sequences for *C. cryptanthus* from
two geographically distinct locations as well as those of an albino individual and
several photosynthetic relatives. I used these data to investigate plastome diversity
within a mycoheterotrophic species as well as to examine the extent of plastome
degradation – with respect to functional loss and structural change – across the
genus. Chapter 3 describes the plastome sequence of *D. australis*. Since *Danhatchia*
is taxonomically isolated I compare this sequence to publically available sequences
for other members of the subtribe Goodyerinae. Finally, Chapter 4 compares results
for the two genera and discusses the evolutionary implications of these analyses.

1.7 References

Abadie, J. C., Püttsepp, Ü., Gebauer, G., Faccio, A., Bonfante, P., & Selosse, M. A. (2006).
    Cephalanthera longifolia (Neottieae, Orchidaceae) is mixotrophic: a comparative
    study between green and nonphotosynthetic individuals. *Canadian Journal of
    Botany, 84*, 1462-1477.
Allen, J. F. (1993). Control of gene expression by redox potential and the requirement for
    chloroplast and mitochondrial genomes. *Journal of Theoretical Biology, 165*, 609-
    631.

Allen, J. F. (2003). Why chloroplasts and mitochondria contain genomes. *Comparative and Functional Genomics, 4*, 31-36.

Banks, D. (2012). *Danhatchia australis* in Australia. *Australian Orchid Review, 77*, 2.

Barrett, C. F., & Davis, J. I. (2012). The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany, 99*, 1513-1523.

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., & Santos, C. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution, 31*, 3095-3112.

Bélanger, A. S., Brouard, J. S., Charlebois, P., Otis, C., Lemieux, C., & Turmel, M. (2006). Distinctive architecture of the chloroplast genome in the chlorophycean green alga *Stigeoclonium helveticum. Molecular Genetics and Genomics, 276*, 464-477.

Bellot, S., & Renner, S. S. (2016). The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution, 8*, 189-201.

Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M. L., & Ruhlman, T. A. (2016). Variable presence of the inverted repeat and plastome stability in *Erodium. Annals of Botany, 117*, 1209-1220.

Breitwieser, I., Brownsey, P. J., Garnock-Jones, P. J., Perrie, L. R., & Wilton, A. D. (2012). Phylum Tracheophyta: vascular plants. In D. P. Gordon (Ed.), *New Zealand inventory of biodiversity.* (pp. 411–459). Christchurch: Canterbury University Press.

Brundrett, M. C. (2002). Coevolution of roots and mycorrhizas of land plants. *New Phytologist, 154*, 275-304.

Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics, 44*, 281-302.

Chang, C. C., Lin, H. C., Lin, I. P., Chow, T. Y., Chen, H. H., Chen, W. H., Cheng, C. H., Lin, C. Y., Liu, S. M., Chang, C. C., & Chaw, S. M. (2006). The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Molecular Biology and Evolution, 23*, 279-291.

Daley, D. O., & Whelan, J. (2005). Why genes persist in organelle genomes. *Genome Biology, 6*, 110-110.

Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology, 17*, 134.

Dawson, M. I., Molloy, B. P. J., & Beuzenberg, E. J. (2007). Contributions to a chromosome atlas of the New Zealand flora—39. Orchidaceae. *New Zealand Journal of Botany, 45*, 611-684.

de Cambiaire, J. C., Otis, C., Lemieux, C., & Turmel, M. (2006). The complete chloroplast genome sequence of the chlorophycean green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. *BMC Evolutionary Biology, 6*, 37.

Delannoy, E., Fujii, S., Colas Des Francs-Small, C., Brundrett, M., & Small, I. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution, 28*, 2077-2086.

Dyall, S. D., Brown, M. T., & Johnson, P. J. (2004). Ancient invasions: From endosymbionts to organelles. *Science, 304*, 253-257.

Ellis, R. J. (1970). Further similarities between chloroplast and bacterial ribosomes. *Planta, 91*, 329-335.

Endo, T., Shikanai, T., Takabayashi, A., Asada, K., & Sato, F. (1999). The role of chloroplastic NAD(P)H dehydrogenase in photoprotection. *FEBS Letters, 457*, 5-8.

Fairman, J. W., Noinaj, N., & Buchanan, S. K. (2011). The structural biology of β-barrel membrane proteins: a summary of recent reports. *Current Opinion in Structural Biology, 21*, 523-531.

Feng, Y. L., Wicke, S., Li, J. W., Han, Y., Lin, C. S., Li, D. Z., Zhou, T. T., Huang, W. C., Huang, L. Q., & Jin, X. H. (2016). Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biology and Evolution, 8*, 2164-2175.

Graham, S. W., Lam, V. K. Y., & Merckx, V. S. F. T. (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist, 214*, 48-55.

Hatch, E. D. (1956). A new species of *Corybas* Salisbury, and a note on some name changes in *Wahlenbergia* Schrader. *Transactions of the Royal Society of New Zealand, Botany 83*, 577.

Inaba, T., & Ito-Inaba, Y. (2010). Versatile roles of plastids in plant growth and development *Plant and Cell Physiology, 51*, 6.

Jones, D. L., & Clements, M. A. (2018). *Danhatchia novaehollandiae* (Orchidaceae: Goodyerinae), a new species from south-eastern Australia. *Australian Orchid Review, 84*, 2.

Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., & Selosse, M. A. (2005). Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytologist, 166*, 639-653.

Kim, H. T., & Chase, M. W. (2017). Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). *PloS ONE, 12*, e0187318-e0187318.

Kim, H. T., Shin, C. H., Sun, H., & Kim, J. H. (2018). Sequencing of the plastome in the leafless green mycoheterotroph *Cymbidium macrorhizon* helps us to understand an early stage of fully mycoheterotrophic plastome structure. *Plant Systematics and Evolution, 304*, 245-258.

Kleine, T., Maier, U. G., & Leister, D. (2009). DNA transfer from organelles to the nucleus: The idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology, 60*, 115-138.

Lavin, M., Doyle, J. J., & Palmer, J. D. (1990). Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution, 44*, 390-402.

Leake, J. R. (1994). The biology of myco-heterotrophic ('saprophytic') plants. *New Phytologist, 127*, 171-216.

Lehnebach, C. A., Rolfe, J. R., Gibbins, J., & Ritchie, P. (2016a). Two new species of *Gastrodia* (Gastrodieae, Orchidaceae) endemic to New Zealand. *Phytotaxa, 277*, 237-253.

Lehnebach, C. A., Zeller, A. J., Frericks, J., & Ritchie, P. (2016b). Five new species of *Corybas* (Diurideae, Orchidaceae) endemic to New Zealand and phylogeny of the Nematoceras clade. *Phytotaxa, 270*, 1-24.

Leonid, A. S., Paul, A. B., & Peter, J. N. (1998). The plastid *ndh* genes code for an NADH-specific dehydrogenase: Isolation of a complex I analogue from pea thylakoid membranes *Proceedings of the National Academy of Sciences of the United States of America, 95*, 1319-1324.

Levin, D. A. (2003). The cytoplasmic factor in plant speciation. *Systematic Botany, 28*, 5-11.

Lim, G. S., Barrett, C., Pang, C. C., & Davis, J. I. (2016). Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *American Journal of Botany, 103*, 1129-1137.

Logacheva, M. D., Schelkunov, M. I., Nuraliev, M. S., Samigullin, T. H., & Penin, A. A. (2014). The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution, 6*, 238-246.

Logacheva, M. D., Schelkunov, M. I., & Penin, A. A. (2011). Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis. Genome Biology and Evolution, 3*, 1296-1303.

Logacheva, M. D., Schelkunov, M. I., Shtratnikova, V. Y., Matveeva, M. V., & Penin, A. A. (2016). Comparative analysis of plastid genomes of non-photosynthetic Ericaceae and their photosynthetic relatives. *Scientific Reports, 6*, 30042.

Lohse, M., Drechsel, O., Kahlau, S., & Bock, R. (2013). OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research, 41*, 575-581.

Lyon, P. S. (2015). *Molecular systematic, biogeography, and mycorrhizal associations in the Acianthinae (Orchidaceae), with a focus on the genus Corybas.* (PhD Thesis), University of Wisconsin-Madison, USA.

Manuell, A. L., Quispe, J., & Mayfield, S. P. (2007). Structure of the chloroplast ribosome: novel domains for translation regulation. *PLoS biology, 5*, e209-e209.

Maréchal, A., & Brisson, N. (2010). Recombination and the maintenance of plant organelle genome stability. *New Phytologist, 186*, 299-317.

Margolin, W. (2005). FtsZ and the division of prokaryotic cells and organelles. *Nature Reviews. Molecular Cell Biology, 6*, 862-871.

Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven, CT: Yale University Press.

Martin, W., & Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiology, 118*, 9-17.

Martin, W., & Kowallik, K. (1999). Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. *European Journal of Phycology, 34*, 287-295.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., & Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences, 99*, 12246-12251.

McCoy, S. R., Kuehl, J. V., Boore, J. L., & Raubeson, L. A. (2008). The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evolutionary Biology, 8*, 130.

Merckx, V. (2012). *Mycoheterotrophy: The Biology of Plants Living on Fungi*. New York: Springer-Verlag.

Merckx, V., Bidartondo, M. I., & Hynson, N. A. (2009). Myco-heterotrophy: when fungi host plants. *Annals of Botany, 104*, 1255-1261.

Miyagishima, S. Y. (2011). Mechanism of plastid division: From a bacterium to an organelle. *Plant Physiology, 155*, 1533-1544.

Moctezuma, E., & Feldman, L. J. (1999). The role of amyloplasts during gravity perception in gynophores of the peanut plant (*Arachis hypogaea*). *Annals of Botany 84*, 709-713.

Molina, J., Hazzouri, K. M., Nickrent, D., Geisler, M., Meyer, R. S., Pentony, M. M., Flowers, J. M., Pelser, P., Barcelona, J., Inovejas, S. A., Uy, I., Yuan, W., Wilkins, O., Michel, C. I., LockLear, S., Concepcion, G. P., & Purugganan, M. D. (2014). Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution, 31*, 793-803.

Moore, L. B., & Edgar, E. (1970). *Indigenous Tracheophyta* (Vol. II). Wellington, New Zealand: Government Printer.

Mower, J. P., & Vickrey, T. L. (2018). Chapter Nine - Structural diversity among plastid genomes of land plants. In S. M. Chaw & R. K. Jansen (Eds.), *Advances in Botanical Research* (Vol. 85, pp. 263-292): Academic Press.

Nickrent, D. L. (2002). Parasitic Plants of the World. In J. A. López-Sáez, P. Catalán, & L. Sáez (Eds.), Parasitic Plants of the Iberian Peninsula and Balearic Islands (pp. 7-27). S. A., Madrid: Mundi-Prensa Libros

Ogura-Tsujita, Y., Yokoyama, J., Miyoshi, K., & Yukawa, T. (2012). Shifts in mycorrhizal fungi during the evolution of autotrophy to mycoheterotrophy in *Cymbidium* (Orchidaceae). *American Journal of Botany, 99*, 1158-1176.

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annual Review of Genetics, 19*, 325-354.

Palmer, J. D., & Thompson, W. F. (1981). Rearrangements in the chloroplast genomes of mung bean and pea. *Proceedings of the National Academy of Sciences of the United States of America, 78*, 5533-5536.

Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adamska, I., & van Wijk, K. J. (2000). Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell, 12*, 319-341.

Perry, A. S., & Wolfe, K. H. (2002). Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution, 55*, 501-508.

Pfannschmidt, T. (2003). Chloroplast redox signals: how photosynthesis controls its own genes. *Trends in Plant Science, 8*, 33-41.

Press, M. C., Scholes, J. D., & Watling, J. R. (1999). Parasitic plants: physiological and ecological interactions with their hosts. In M. C. Press, Scholes, J. D., Barker, M. G., (Ed.), *Physiological Plant Ecology* (pp. 175–197). Oxford, UK: Blackwell Science.

Pyke, K. A. (1999). Plastid division and development. *Plant Cell, 11*, 549-556.

Ravin, N. V., Gruzdev, E. V., Beletsky, A. V., Mazur, A. M., Prokhortchouk, E. B., Filyushin, M. A., Kochieva, E. Z., Kadnikov, V. V., Mardanov, A. V., & Skryabin, K. G. (2016). The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC plant biology, 16*, 153–161.

Rowan, B. A., Oldenburg, D. J., & Bendich, A. J. (2010). RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *Journal of Experimental Botany, 61*, 2575-2588.

Sanderson, M. J., Copetti, D., Burquez, A., Bustamante, E., Charboneau, J. L., Eguiarte, L. E., Kumar, S., Lee, H. O., Lee, J., McMahon, M., Steele, K., Wing, R., Yang, T. J., Zwickl, D., & Wojciechowski, M. F. (2015). Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany, 102*, 1115-1127.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., & Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research, 6*, 283-290.

Schelkunov, M. I., Shtratnikova, V. Y., Nuraliev, M. S., Selosse, M. A., Penin, A. A., & Logacheva, M. D. (2015). Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution, 7*, 1179-1191.

Smith, D. R., & Lee, R. W. (2014). A plastid without a genome: Evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiology, 164*, 1812-1819.

St. George, I. (1999). *The Natural Guide to New Zealand Orchids*. New Zealand: Random House.

Steenbeeke, G. (2012). *Danhatchia australis* - not just a New Zealand species. *Orchadian, 17*, 258.

Strauss, S. H., Palmer, J. D., Howe, G. T., & Doerksen, A. H. (1988). Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proceedings of the National Academy of Sciences of the United States of America, 85*, 3898-3902.

Suetsugu, K., Ohta, T., & Tayasu, I. (2018). Partial mycoheterotrophy in the leafless orchid *Cymbidium macrorhizon*. *American Journal of Botany, 105*, 1595-1600.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics, 5*, 123-135.

Turmel, M., Otis, C., & Lemieux, C. (2005). The complete chloroplast DNA sequences of the charophycean green algae *Staurastrum* and *Zygnema* reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. *BMC Biology, 3*, 22.

Turmel, M., Otis, C., & Lemieux, C. (2006). The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Molecular Biology and Evolution, 23*, 1324-1338.

Turmel, M., Pombert, J. F., Charlebois, P., Otis, C., & Lemieux, C. (2007). The green algal ancestry of land plants as revealed by the chloroplast genome. *International Journal of Plant Sciences, 168*, 679-689.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., & Sugiura, M. (1994). Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii. Proceedings of the National Academy of Sciences of the United States of America, 91*, 9794-9797.

Whatley, J. M., & Whatley, F. R. (1981). Chloroplast evolution. *New Phytololgist, 87*, 233-247.

Wicke, S., Müller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., Renner, S. S., & Schneeweiss, G. M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell, 25*, 3711-3725.

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America, 84*, 9054-9058.

Wolfe, K. H., Morden, C. W., & Palmer, J. D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 10648-10652.

Xu, J. H., Liu, Q., Hu, W., Wang, T., Xue, Q., & Messing, J. (2015). Dynamics of chloroplast genomes in green plants. *Genomics, 106*, 221-231.

Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., Wang, D., Lai, C., Xu, W., Huang, J., Zha, L., Liu, D., Ma, X., Wang, L., Zhou, M., Jiang, Z., Meng, H., Peng, H., Liang, Y., Li, R., Jiang, C., Zhao, Y., Nan, T., Jin, Y., Zhan, Z., Yang, J., Jiang, W., & Huang, L. (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nature Communications, 9*, 1615-1615.

Chapter 2

The complete chloroplast genomes of photosynthetic and non-photosynthetic *Corybas* (Orchidaceae): tracing plastome evolution across a transition to mycoheterotrophy

2.1 Introduction

Corybas Salisb. (Diurideae, Orchidaceae) is a genus of approximately 135 small, terrestrial orchids distributed across Asia, Australia, New Zealand and the Pacific Islands (Lyon, 2015; Merckx, 2012). About 40 species are native to New Guinea, with the remainder mostly occurring in Australia and New Zealand (Lyon, 2015). High rates of endemism are a feature of *Corybas* (Lyon, 2015) and most of the approximately 20 New Zealand species are thought to be endemic (Breitwieser et al., 2019). Members of *Corybas* are generally perennial herbs producing a single, orbiculate-cordate leaf that lies close to the ground and a single complex, highly modified flower (Breitwieser et al., 2019; Dransfield et al., 1986; Lyon, 2015; Moore & Edgar, 1970). Growth is seasonal; tubers persist below ground during dry or cold periods with leaves and flowers emerging above ground when conditions are favourable (Moore & Edgar, 1970; Watkins, 2012).

A phylogeny of *Corybas* based on several chloroplast and nuclear markers (Lyon, 2015) recovered eight major clades and identified several species that did not clearly fall within any of these groups. Four clades comprise Australian and New Zealand species, with the remainder comprising species from mainland Asia, New Guinea, and Sundaland. The Australian and New Zealand species have been studied taxonomically and several infrageneric groupings are currently recognised. These taxonomic groups correspond closely to clades identified by Lyon (2015). Two of the clades (i.e., Anzybas and Singularybas) contain representatives from both Australia and New Zealand, one contains representatives from Australia (i.e., Corysanthes), and one contains representatives from New Zealand (i.e., Nematoceras). Two New Zealand species fall outside of these clades; *Corybas oblongus* (Hook.f.) Rchb.f., which is sister to the Singularybas clade, and *Corybas cryptanthus* Hatch, which is sister to the pairing of the Corysanthes and Nematoceras clades (Lehnebach et al.,

2016; Lyon, 2015). In the analyses of Lyon (2015), most of the New Zealand *Corybas* belong to the Nematoceras clade (e.g., *C. iridescens* Irwin and Molloy, *C. macranthus* (Hook. f.) Rchb. f.) with two placed within the Anzybas clade (e.g., *Corybas carsei* (Cheeseman) Hatch) and one within the Singularybas clade (i.e., *Corybas cheesemanii* (Hook. f. ex Kirk) Kuntze).

Bayesian relaxed clock analyses using the results of Gustafsson et al. (2010) as secondary calibrations suggest that *Corybas* and its sister, *Cyrtostylis* R.Br., diverged ~18 million years ago (Lyon, 2015). This observation is consistent with the idea that long distance dispersal has been important in the evolution of New Zealand plants, with many plant lineages arriving in New Zealand after it had become isolated from Gondwana ~85 million years ago (Winkworth et al., 2002). Phylogenetic results for *Corybas* suggest six independent dispersal events between Australia and New Zealand. Four of the events likely involved a single dispersal from Australia to New Zealand. Specifically, two events involving representatives of the Anzybas clade (i.e., *C. carsei* and *C. rotundifolius* (Hook f.) Rchb. f.), one involving the Singularybas clade (i.e., *C. cheesemanii*), and the last involving *C. oblongus*. For the remaining two events, involving *C. cryptanthus* together with the Corysanthes and Nematoceras clades, two scenarios possibly explain the direction of dispersal. One possibility is that the common ancestor of these three groups dispersed from Australia to New Zealand, with the ancestor of the Corysanthes clade subsequently dispersing back to Australia. The other is that the ancestors of *C. cryptanthus* and the Nematoceras clade dispersed independently from Australia to New Zealand. Results for *Corybas* are consistent with those from broader analyses that suggest plant dispersal from Australia to New Zealand is much more common than that from New Zealand to Australia (Sanmartín et al., 2007).

Most of the New Zealand species of *Corybas* belong to the Nematoceras clade. Members of this clade are often morphologically and genetically similar; consistent with the results of Lyon (2015) this suggests that the Nematoceras clade has diversified within the last 5 million years. Within this clade there are a mixture of formally recognised species and as yet undescribed entities. For example, the Trilobus aggregate contains several species that are supported by genetic and morphological evidence (e.g., *Corybas confusus* Lehnebach and *Corybas vitreus* Lehnebach) as well as species and tag name entities (e.g., *Corybas* "Remutaka") that

24

are distinguished morphologically, but not clearly distinct in genetic analyses (Lehnebach et al., 2016).

All but one species of *Corybas* are green and presumably photosynthetic. The exception is the New Zealand endemic *C. cryptanthus*, which is assumed to be non-photosynthetic based on its lack of green pigmentation and lack of leaves on the aboveground portion of the plant (Hatch, 1956; Lyon, 2015; Merckx, 2012; Moore & Edgar, 1970). Anatomical investigations have also indicated that fungi penetrate the tissues of the below ground portion of the plant (Campbell, 1972). Together these observations strongly suggest that *C. cryptanthus* is holomycoheterotrophic.

As is the case for many holomycoheterotrophs, the vegetative body of *C. cryptanthus* remains underground in the form of a rhizome. In this species, the rhizome is up to 5 cm in length and 1 mm in diameter, many branched, and while it bears scale leaves and minute tubers, it lacks roots. Again, typical of a holomycoheterotroph, only the reproductive organs of *C. cryptanthus* appear above ground. Flowers emerge in June to October, and, after pollination, the capsule is raised above the forest floor on a tall peduncle. The flowers are structurally similar to those of other *Corybas*, but are translucent and the labellum more coarsely toothed than in other New Zealand species (de Lange, 2005; Moore & Edgar, 1970). *Corybas cryptanthus* occurs throughout the North and South Islands, as well as on the Three Kings Island, typically in dense shrub or forest vegetation. The species is often found under kanuka (*Kunzea* Rchb. pp., Myrtaceae) or beech (*Nothofagus* Blume. pp., Nothofagaceae) and often alongside *C. cheesemanii* (Campbell, 1972; de Lange, 2005). Specimens of *C. cryptanthus* are often overlooked, probably due to their short stature, growth habit, and translucent flower (Irwin, 1954).

Although only a single non-photosynthetic species is currently recognised within *Corybas*, albino and variegated (partially albino) individuals have been found at Waitarere Forest (Horowhenua, New Zealand) (Figure 2.1). This is the first report of albino individuals for *Corybas*, although albinos have been recorded from other orchid genera, including *Cephalanthera* Rich. and *Epipactis* Zinn (Abadie et al., 2006; Julou et al., 2005; Salmia, 1989; Selosse et al., 2004). At the Waitarere site, albino and variegated plants are scattered amongst large populations of photosynthetic *Corybas*, which are largely members of the Trilobus aggregate (e.g., *Corybas obscurus* Lehnebach) (Figure 2.1).

25

Figure 2.1 Albino, variegated (partially albino), and photosynthetic *Corybas* Salisb. from Waitarere (Horowhenua, New Zealand). A, an albino individual amongst green *Corybas* congeners (photograph taken 7 October 2017). B and C, variegated *Corybas* individuals amongst green *Corybas* congeners (photographs taken 2 September 2018).

Due to the apparent lack of chlorophyll in some or all aboveground portions of these individuals, their photosynthetic capacity is likely to be compromised and they are, therefore, likely to rely more heavily on their fungal associates. Albino and variegated plants were observed during visits to the site in both 2017 and 2018. However, it is unclear whether these are the result of long-term (e.g., year to year) persistence or the repeated origin of short-lived (i.e., one growing season) albinos.

Currently, complete plastome sequences are available for 32 recognised species of mycoheterotrophs. In most cases a single accession of each species has been sequenced and often plastomes from photosynthetic relatives are unavailable. This limits the inferences that can be made about the patterns of plastome evolution associated with the transition to mycoheterotrophy. For example, the loss of the NADH dehydrogenase subunit (*ndh*) genes has recently been linked to

destabilisation of the IR boundaries in orchids (Kim et al., 2015; Niu et al., 2017) as well as to the initial stages of plastome degradation in mycoheterotrophs (Barrett & Davis, 2012; Barrett et al., 2014). Comparative analyses of plastome sequences from mycoheterotrophs and close photosynthetic relatives are needed to rigorously test these ideas.

The present study describes the assembly of complete plastome sequences for 11 representatives of *Corybas* plus those for *Microtis unifolia* (G.Forst.) Rchb.f. and *Chiloglottis cornuta* Hook.f., two photosynthetic Diurideae native to New Zealand. Sampling of *Corybas* includes seven photosynthetic species from New Zealand and one from Australia, as well as two geographically distinct populations of the holomycoheterotrophic *C. cryptanthus*, and that of an albino individual from Waitarere. Comparative analyses of these plastome sequences indicate marked variation in genome structure across *Corybas* and extensive plastome reduction in *C. cryptanthus* but not in the albino. Plastomes from the two *C. cryptanthus* populations are also found to be similar. These results are discussed in the context of the Barret and Davis' model of plastome evolution in mycoheterotrophic plants (Barrett & Davis, 2012; Barrett et al., 2014)

2.2 Materials and Methods

2.2.1 Taxon sampling
Tissue samples were sourced from existing, herbarium-vouchered collections or newly collected plant material. Details of the collections are provided in Table 2.1.

2.2.2 Extraction of total cellular DNA
DNA extractions were conducted using NucleoSpin Plant II kits (Macherey-Nagel) following the manufacturer's instructions. Once extracted, DNA concentrations were quantified using a Qubit 2.0 Fluorimeter (LifeTechnologies) according to the manufacturer's instructions.

2.2.3 Illumina short read sequencing
The Illumina TruSEQ DNA Nano library preparation methods were used to prepare *Corybas, M. unifolia* and *Ch. cornuta* libraries. Briefly, TruSEQ DNA Nano library preparation involves shearing the DNA by ultrasonication into random fragments,

Table 2.1 Collection details for the included *Corybas* and related taxa

Species and authority	Locality	Voucher
Chiloglottis cornuta Hook.f.	QEII Reserve, Wellington (George Gibbs)	WELT SP107841
Corybas cheesemanii (Hook.f. ex Kirk) Kuntze	QEII Reserve, Wellington (George Gibbs)	WELT SP107840
Corybas cryptanthus Hatch	Omoana, Taranaki	WELT SP107693
Corybas cryptanthus Hatch	East Harbour Regional Park, Wellington	WELT SP107836
Corybas diemenicus (Lindl) Rupp	Tasmania, Australia	Chase O-564
Corybas dienemus D.L. Jones	Tararua Forest Park, Wellington	WELT SP104218
Corybas iridescens Irwin & Molloy	Atene, Whanganui	WELT SP107705
Corybas macranthus (Hook.f.) Rchb.f.	East Harbour Regional Park, Wellington	WELT SP107834
Corybas obscurus Lehnebach	Nelson Lakes National Park, Nelson	WELT SP104405
Corybas "Remutaka"	QEII Reserve, Wellington (George Gibbs)	WELT SP107839
Corybas vitreus Lehnebach	Mt. Richmond Forest Park, Nelson-Marlborough	WELT SP105876
Corybas "Waitarere white"	Waitarere, Horowhenua	No voucher
Microtis unifolia (G.Forst.) Rchb.f	Hataitai, Wellington	WELT SP107835

with barcoded adapters added to each end during PCR enrichment. The Massey Genome Service (MGS) performed the library preparation and paired-end DNA sequencing on Illumina MiSeq instruments. Following sequencing, the sequence reads were quality assessed using a standard SolexaQA-based workflow (Cox et al., 2010).

2.2.4 Chloroplast genome assembly

Whole chloroplast genome sequences were assembled from quality assessed Illumina 250 bp sequence reads using a genome skimming approach. Next-generation sequencing of total cellular DNA produces reads from all three plant genomes – nuclear, chloroplast and mitochondrial. Genome skimming allows reads from the chloroplast genome to be isolated from this mixture of reads. The chloroplast genome can therefore be assembled without isolating the chloroplast DNA prior to sequencing (Straub et al., 2012; Twyford & Ness, 2017). The custom bioinformatics pipeline used here combined idba_ud v. 1 (Peng et al., 2012), BLAST + 2.3 (Altschul et al., 1990), Geneious R9 (Kearse et al., 2012) and BWA v. 0.7 (Li & Durbin, 2009), as described below.

First, quality assessed reads were assembled into contigs using idba_ud, an iterative de Bruijn graph assembler for paired-end reads that is particularly useful when coverage of the genome is likely to be uneven (Peng et al., 2012). The resulting contigs were then filtered to remove those that had low similarity to a reference

collection of 70 publicly available, complete Orchidaceae plastome sequences that included both photosynthetic and non-photosynthetic species (Appendix I). Custom scripts were used to first remove contigs shorter than 150 nt from the idba_ud results and then to automate BLAST searches against the reference collection. Filtered contigs were then assembled into larger scaffolds using the reference-based and *de novo* assembly tools implemented in Geneious. Reference-based assemblies used the photosynthetic *Goodyera procera* (Ker Gawl.) Hook. (Yu, Liang, & Wu, 2015) and *Sobralia callosa* L.O.Williams (Kim et al., 2015) genomes.

The IR boundaries were located by first identifying contigs and scaffolds that contained genes generally located at these boundaries (e.g., ribosomal protein subunit 22 (*rpl22*) and hypothetical chloroplast RF1 (*ycf1*)). Typically, contigs and scaffolds in this set were composed of two distinct sequence portions; one with high identity to all members of this set, the other with high identity to only some members of the set. The IR boundaries were assumed to occur at the point at which identity shifted between these two situations. To complete the circular plastome, a second copy of the IR was then copied into the sequence manually.

As a final check, the original Illumina short sequence reads were mapped directly to the genome drafts using BWA and visualised in Tablet v. 1.16 (Milne et al., 2010). Inconsistencies in read coverage were assumed to indicate assembly errors and drafts were modified as necessary to correct these.

2.2.5 Chloroplast genome annotation

Draft genomes were annotated using tools implemented in Geneious R9 with the *S. callosa* and *G. procera* plastomes as references. To capture the largest number of putative gene regions, initial searches used a low similarity threshold (i.e., 35%).

For putative protein coding genes, the coding sequence was translated and the expected product inspected by eye for appropriate start and stop codons, as well as overall similarity in terms of both length and amino acid sequence to existing sequences. Each of the putative protein coding regions was then classified as either a functional gene, a pseudogene, or a gene remnant based upon similarity to existing, putatively functional gene sequences.

A putative protein coding region was considered functional if, compared to presumably functional copies of the corresponding gene from other species, the sequence differed by <10% in length and the amino acid positive identity, a

description of the extent to which amino acid substitutions disrupt the physiochemical properties (e.g., charge or hydrophobicity) of a protein sequence, was 75% based on the BLOSUM62 matrix. Additionally, where functional domains or residues had been identified these also needed to be retained for the gene to be considered a functional copy. A putative coding region was classed as a pseudogene if it remained recognisable as the corresponding gene, but the amino acid translation suggested it was unlikely to be functional (e.g., sequence was truncated or missing functional domains or residues). Specifically, a putative coding sequence from Diurideae was assumed to be a pseudogene when the sequence was considerably shorter than presumably functional copies of the corresponding gene (i.e., 21-40% shorter), but retained high amino acid identity (i.e., >90%) or when the sequence was more complete in terms of length (i.e., 11-20% shorter) but amino acid identity was lower (i.e., 60-80%). Finally, a putative coding sequence was considered a gene remnant when the sequence was short (i.e., 41-61% shorter) but had high identity (i.e., >80%) with presumably functional copies of a gene, or when the sequence was longer (i.e., 0-40% shorter) but had lower identity (i.e., 60-89%).

The low similarity threshold used for the initial search resulted in annotations for many putative transfer RNA (tRNA) and ribosomal RNA (rRNA) genes. Annotations were checked manually and where similarity scores were <80% the annotation was removed.

Comparative analyses

The assembled plastomes were compared for several features. A number of these features, such as plastome size and GC content, were recorded from Geneious. Other features were calculated independently.

Repeated sequences — The Geneious repeat finder tool was used to locate repeats in the LSC and SSC regions of each plastome; repeats unique to the IR were excluded from this analysis. A pair of searches were carried out on each plastome. The first identified repeats of > 30 nt in length. This generated few repeats, so the genomes were also annotated for repeats of ≥ 15 nt. Preliminary searches for smaller repeats (10-19 nt in length) consistently resulted in very large numbers per genome (e.g., >10,000) and these were not analysed further. Only exact repeats with 100%

identity were analysed and repeats wholly enclosed within another longer repeat sequence were also excluded.

Single nucleotide polymorphisms and indels — The Geneious variation/SNP finder tool was used to locate single nucleotide polymorphisms (SNPs) and other variations between pairs of *Corybas* plastomes, with the IR and LSC regions considered separately. Whole genomes for photosynthetic *Corybas* were first aligned using LASTZ in Geneious with *Corybas* "Remutaka" as the target sequence. Pairwise analyses between species for the IR and for the LSC regions were then carried out. To deal with differing boundaries between the SC and IR regions, the IR was standardized for each pair. The IR was considered to start at the first gene fully enclosed within the IR in both species (usually the ribosomal protein L32 gene (*rpl32*)) through to, and inclusive of, the ribosomal protein S3 gene (*rps3*).

The whole genomes of the two *C. cryptanthus* accessions were also aligned. The IR analogue was taken as being inclusive of *rps3* through to the beginning of the transfer RNA gene for lysine (*trnK*); includes all the genes, or at least portions of them, found in the IR of photosynthetic *Corybas*. This was 26,610 nt in length for both *C. cryptanthus* accessions. The remainder of the plastome was assumed to be an LSC analogue since none of the genes present in the SSC region of photosynthetic *Corybas* are retained in the *C. cryptanthus* plastome. Pairwise analyses were conducted between the IR and LSC analogues of the two *C. cryptanthus* accessions as well as the IR analogues and the IR regions of the photosynthetic *Corybas*. The LSC analogues of the *C. cryptanthus* accessions were too dissimilar to align with the LSC regions of the photosynthetic taxa. For each pairwise comparison the number of SNPs and indels were counted for each region.

Gene Variation — Using both nucleotide and inferred protein sequences, multiple sequence alignments were constructed for all putatively functional protein-coding genes from the Diurideae plastomes using ClustalO (Sievers et al., 2011). For amino acid sequences that included a premature stop codon, the sequence beyond this point was considered for similarity if the reading frame was otherwise intact and amino acid identity remained high (i.e., >70%). Summary statistics (e.g., length, nucleotide and amino acid identity) were calculated for each gene alignment. Two-sample t-tests (Snedecor & Cochrane, 1989) were used to compare variation in gene

length and identity between genome regions (i.e., LSC, SSC, and IRs) and gene classes based on those identified in the Barrett and Davis (2012) model.

2.3 Results

Whole plastome sequences were assembled for 11 representatives of *Corybas* and two close, photosynthetic relatives, *M. unifolia* and *Ch. cornuta* (Figures 2.2 and 2.3, Appendix II). Summary statistics for the DNA sequencing and plastome assemblies are presented in Table 2.2. Ten of the *Corybas* plastomes were from New Zealand accessions; seven from photosynthetic species or tag-name entities (i.e., *C. cheesemanii*, *Corybas dienemus* D.L.Jones, *C. iridescens*, *C. macranthus*, *C. obscurus*, *C.* "Remutaka", and *C. vitreus*), two from the non-photosynthetic *C. cryptanthus*, and one from an albino individual collected at Waitarere. The remaining plastome was from the photosynthetic Australian species, *Corybas diemenicus* (Lindl.) Rupp. The completed plastome sequences have been deposited in Genbank with the accession numbers MK796239, MK848867-MK848868, MK847775-MK847777, MK876365-MK876368, MK883746-MK883747, MN218689.

2.3.1 Overall genome size and structure
Plastomes from all photosynthetic representatives of *Corybas* are similar in size (Table 2.3). These plastomes differ by only ~1430 nt with the largest being *C. dienemus*, at 134,859 nt in length, and the smallest *C. diemenicus,* at 133,425 nt. Plastomes from the New Zealand species range in size from 134,013 to 134,859 nt; *C. cheesemanii* (Singularybas clade) was the smallest and all others (all from the Nematoceras clade) had overall lengths within 290 nt of the largest, *C. dienemus*. Plastomes for the photosynthetic *Corybas* are also larger than those of *M. unifolia* and *Ch. cornuta*. On average the *Ch. cornuta* plastome was ~5400 nt smaller than those from photosynthetic *Corybas*, whereas that of *M. unifolia* was ~7000 nt smaller.

Plastomes from photosynthetic *Corybas*, *M. unifolia* and *Ch. cornuta* retain the quadripartite plastome structure typical of land plants. Consistent with smaller overall plastome sizes, the IR and LSC regions of *M. unifolia* and *Ch. cornuta* are also 601-8113 nt smaller than those of the photosynthetic *Corybas*. In contrast, the

32
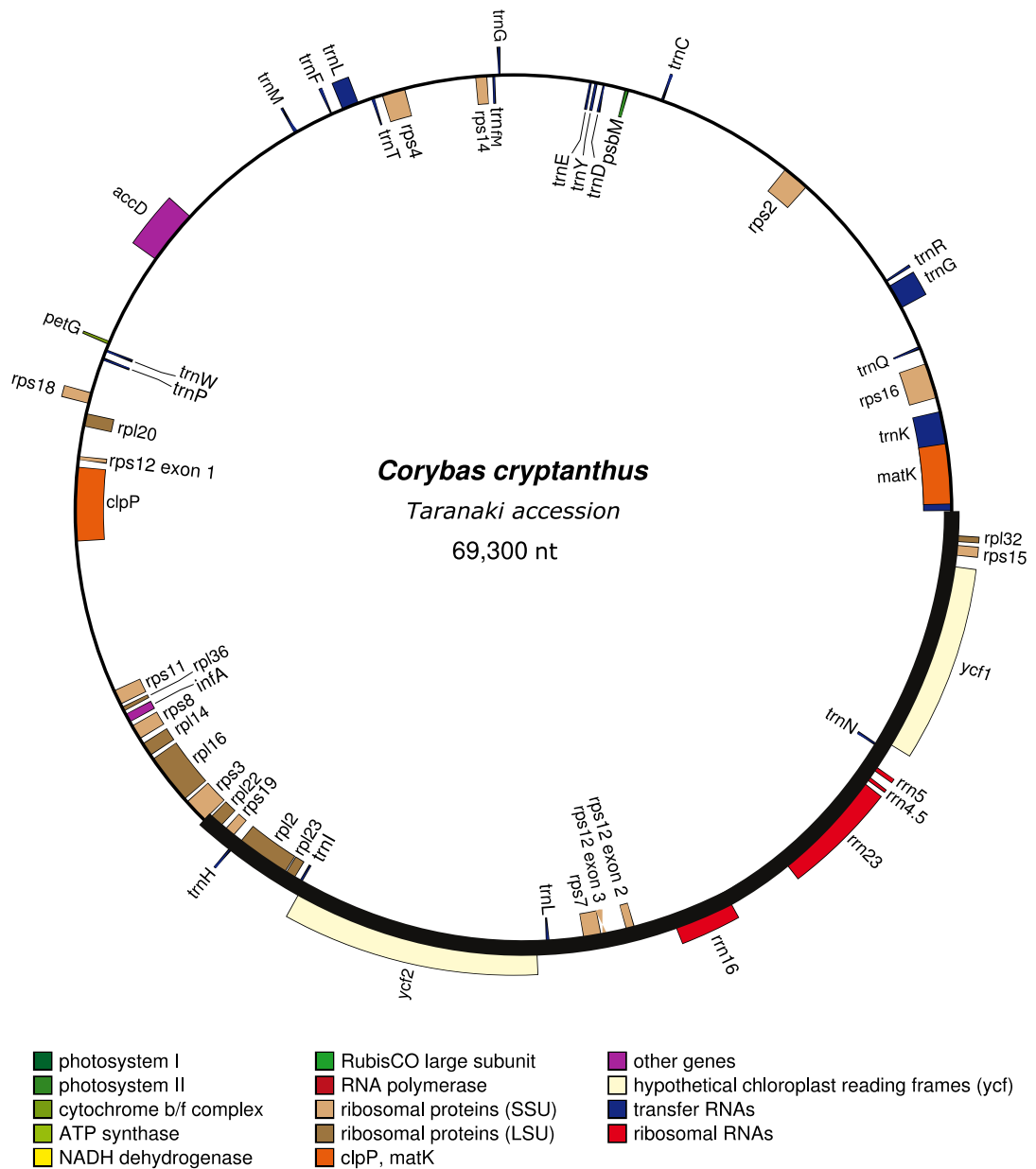
Figure 2.2 The plastome of *Corybas cryptanthus* Hatch Taranaki accession. Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of the circle are transcribed in the reverse direction. The region of the plastome equivalent to the inverted repeat (IR) in photosynthetic *Corybas* (the "IR analogue") is emboldened. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).
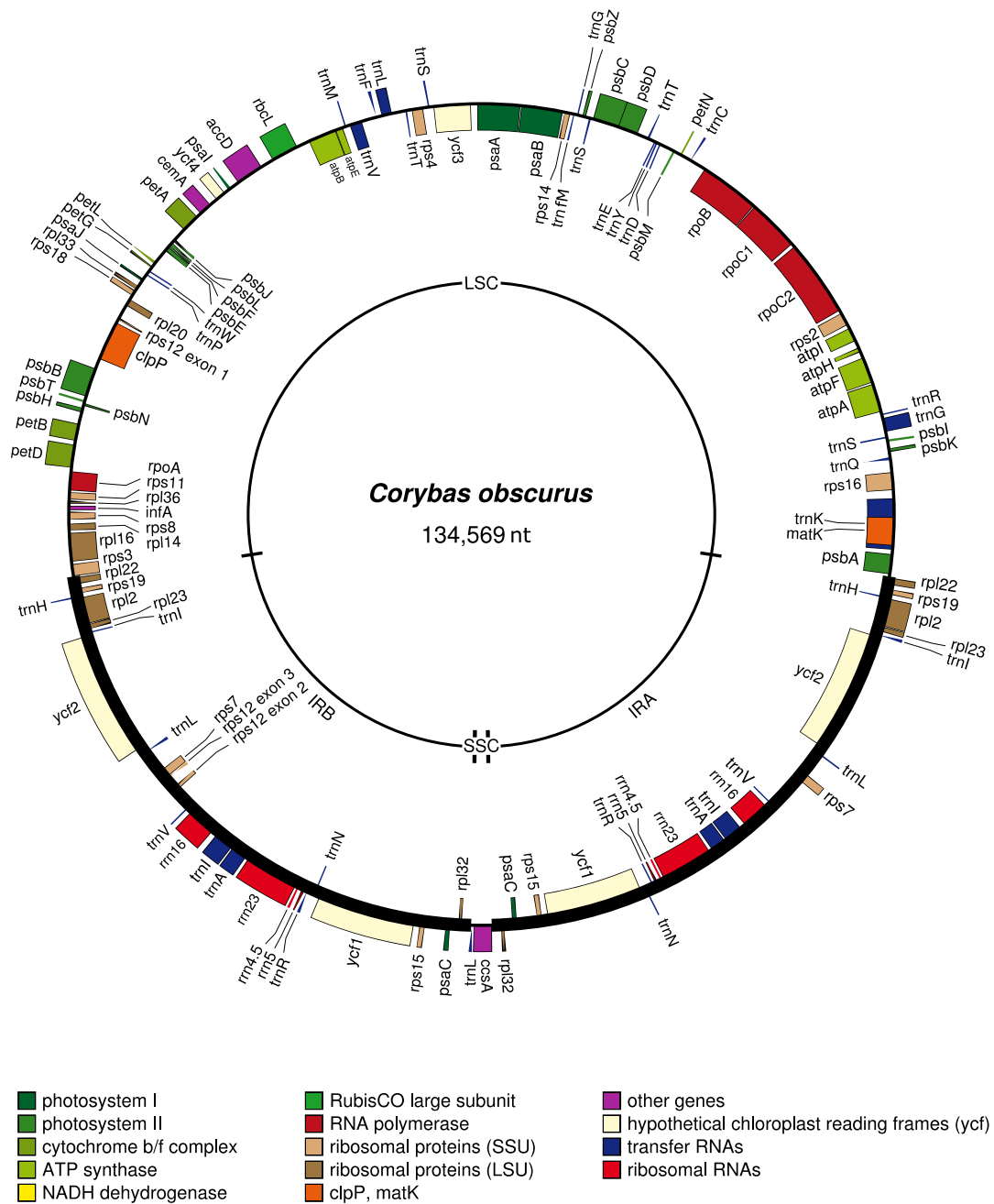
Figure 2.3 The plastome of *Corybas obscurus* Lehnebach. Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR_A and IR_B), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013)

Table 2.2 Summary statistics for Diurideae chloroplast genome assemblies

Species	Number of Illumina short sequence reads	Number of Illumina short sequence reads mapped to plastome	% of reads mapped to plastome	Mean and, in parentheses, range of read coverage			
				Overall	IR	SSC region	LSC region
Chiloglottis cornuta	4,914,888	288,026	5.86	547.9 (208-1015)	567.2 (265-1015)	584.3 (208-977)	528 (220-988)
Corybas cheesemanii	6,420,026	105,520	1.64	195.2 (4-515)	209 (4-504)	102.9 (51-160)	183.9 (26-515)
Corybas cryptanthus[a]	1,770,924	16,443	0.93	69.1 (8-302)	–	–	–
Corybas cryptanthus[b]	4,389,804	15,623	0.36	55.5 (21-84)	–	–	–
Corybas diemenicus	3,112,158	135,266	4.35	104.2 (16-844)	108.7 (20-410)	47.0 (32-62)	102.0 (22-844)
Corybas dienemus	1,958,052	135,200	6.90	181.9 (11-513)	191.5 (11-500)	12 5.6(92-180)	165.2 (12-462)
Corybas iridescens	1,364,000	135,152	9.91	172.5 (20-414)	188.4 (20-411)	164.7 (99-292)	162.2 (20-381)
Corybas macranthus	1,462,420	91,789	6.28	169.5 (11-467)	190.5 (11-467)	203.9 (105-357)	154.6 (15-431)
Corybas obscurus	5,237,978	410,329	7.83	755.9 (385-1119)	739.6 (385-1119)	832.4 (677-1025)	767.2 (435-1043)
Corybas "Remutaka"	6,150,712	139,033	2.26	525.4 (167-999)	567.6 (183-999)	191.3 (167-215)	515.4 (213-974)
Corybas "Waitarere white"	1,748,496	60,662	3.47	109.1 (2-584)	169.4 (5-572)	72.7 (47-103)	59.7 (2-390)
Corybas vitreus	1,205,236	58,672	4.87	108.0 (49-180)	110 (53-180)	126.6 (94-147)	106.1 (49-180)
Microtis unifolia	5,198,358	487,113	9.37	943.4 (483-1246)	951.2 (691-1246)	819.8 (483-1020)	1003.6(535-1186)

[a] This accession was collected in East Harbour Regional Park, Wellington

[b] This accession was collected in Omoana, Taranaki

35

Table 2.3 Summary statistics for the plastome sequences

Species	Plastome		IR			SSC			LSC		
	Size (nt)	GC content (%)	Size (nt)	% plastome	GC content (%)	Size (nt)	% plastome	GC content (%)	Size (nt)	% plastome	GC content (%)
Chiloglottis cornuta	129,103	37.50	23,621	36.59	43.10	8719	6.75	27.40	73142	56.65	35.10
Corybas cheesemanii	134,013	38.20	29,219	43.61	40.80	1310	0.98	31.90	74265	55.52	36.20
Corybas cryptanthus[a]	69376	36.60	26,610[c]	38.36	39.90	–	–	–	42766[d]	61.62	34.50
Corybas cryptanthus[b]	69300	36.60	26,610[c]	38.40	39.90	–	–	–	42690[d]	61.60	34.50
Corybas diemenicus	133,425	38.00	29,070	43.58	40.80	1282	0.96	30.40	74003	55.56	36.00
Corybas dienemus	134,859	38.00	29,910	44.36	40.30	1091	0.81	32.60	73948	54.83	36.10
Corybas iridescens	134,785	38.00	29,932	44.41	40.30	1091	0.81	32.60	73830	54.78	36.10
Corybas macranthus	134,625	38.00	29,905	44.43	40.40	1072	0.80	33.20	73743	54.78	36.10
Corybas obscurus	134,569	38.00	29,866	44.39	40.40	1072	0.80	33.00	73765	54.82	36.10
Corybas "Remutaka"	134,677	38.00	29,904	44.41	40.40	1072	0.80	32.90	73797	54.80	36.10
Corybas vitreus	134,734	38.00	29,866	44.33	40.40	1049	0.78	32.50	73953	54.89	36.10
Corybas "Waitarere white"	134,610	38.00	29,856	44.36	40.40	1072	0.80	33.00	73826	54.84	36.10
Microtis unifolia	127,413	37.90	21,819	34.25	43.90	10,856	8.52	28.90	72919	57.23	35.70

[a] This accession was collected in East Harbour Regional Park, Wellington

[b] This accession was collected in Omoana, Taranaki

[c] IR analogue

[d] LSC analogue

36

SSC regions of these two species are 7409-9807 nt larger (Table 2.2). In the Diurideae plastomes, the combined single copy regions account for 55.58-66.34% of the sequence length and the two copies of the IR account for 34.25-44.43%. The IR occupies more of the plastome in photosynthetic representatives of *Corybas* than in either *Ch. cornuta* or *M. unifolia*; in contrast the SSC region occupies markedly less and the LSC region approximately the same proportion (Table 2.3). A similar pattern continues among the photosynthetic *Corybas*. The *Corybas* species with the smallest plastomes overall (i.e., *C. diemenicus* and *C. cheesemanii*) also have the shortest IR and the largest SSC regions (Table 2.2).

The boundaries between the IRs and LSC regions are similar in the photosynthetic *Corybas*, *M. unifolia* and *Ch. cornuta*. In these plastomes the boundary with IR$_B$ is either within the *rps3* gene or in the intergenic spacer between this gene and the *rpl22* gene. The boundary between the LSC region and IR$_A$ always falls within the intergenic spacer between *rpl22* and the photosystem II protein D1 (*psbA*) gene. The size of this spacer and the exact location of the IR boundary within it differ between species (Figure 2.4). The boundaries between the IRs and SSC regions are much less consistent. The SSC regions of *Ch. cornuta* and *M. unifolia* contain more genes than those of the photosynthetic *Corybas* (Figure 2.4). Specifically, complete coding sequence for 5-6 genes in *Ch. cornuta* and *M. unifolia* compared to 0-2 genes in the photosynthetic *Corybas* are found in the SSC region. In *Ch. cornuta*, the boundary between IR$_A$ and the SSC region is within the *ycf1* gene and in *M. unifolia* it is within the intergenic spacer between *ycf1* and *rpl32*. In contrast, three different boundary positions were identified within the sequenced *Corybas* plastomes.

In most of the accessions, the IR$_A$-SSC boundary is within the transfer RNA gene for leucine (*trnL*) with the IR$_B$-SSC boundary within the intergenic spacer between the cytochrome c biogenesis (*ccsA*) and *rpl32* genes. The exceptions involve *C. cheesemanii*, *C. diemenicus*, and *C. vitreus*. In *C. cheesemanii* the IR$_A$-SSC boundary is in the intergenic spacer between *trnL* and photosystem I iron-sulfur center (*psaC*) genes and the IR$_B$-SSC boundary is within the *rpl32* gene. In *C. diemenicus* the IR$_A$-SSC boundary is in the intergenic spacer between the *rpl32* and *psaC* genes and the IR$_B$-SSC boundary is within the *ccsA*. In *C. vitreus* the IR$_A$-SSC boundary is in the *trnL* gene and the IR$_B$-SSC boundary is within the *ccsA* gene (Figure 2.4).
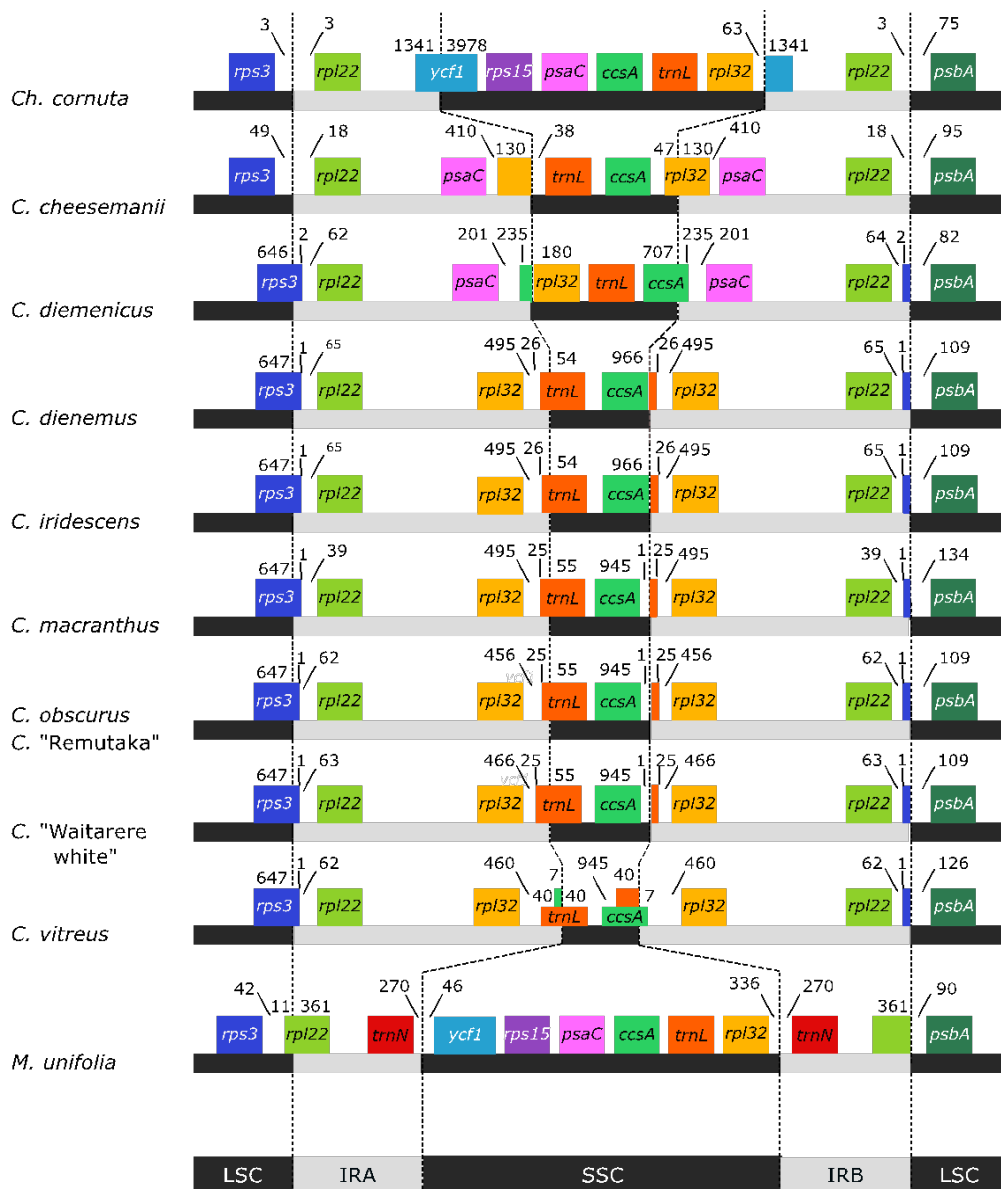
Figure 2.4 Positions of the boundaries between the IR and SC regions in Diurideae plastomes. The boundaries between the LSC region and IR are well-conserved. The LSC-IR$_A$ boundary is always within, or near the terminus, of the ribosomal protein S3 gene (*rps3*), while the LSC-IR$_B$ boundary is always near the start of the photosystem II protein D1 gene (*psbA*). The SSC-IR$_A$ boundary varies between *Corybas* and the other Diurideae; the SSC is larger and contains more genes in *Microtis unifolia* (G.Forst.) Rchb.f. and *Chiloglottis cornuta* Hook.f.

Although both are presumed to be non-photosynthetic, the plastome sequences of the achlorophyllous *C. cryptanthus* and *C.* "Waitarere white" differ markedly in their overall size and structure. The plastome of *C.* "Waitarere white" is typical of the photosynthetic *Corybas*; a quadripartite structure with IR, LSC and SSC regions corresponding to the majority of the photosynthetic *Corybas* accessions (Table 2.2). Plastome sequences for the two *C. cryptanthus* accessions are highly similar to one another but differ markedly from those of the photosynthetic Diurideae. Specifically, the *C. cryptanthus* plastomes are approximately half the size of those from photosynthetic *Corybas* and lack one copy of the IR (Table 2.3; Figures 2.2 and 2.3).

Although these plastomes do not have a quadripartite structure, based on conserved gene content, gene order, and nucleotide similarity, analogues of the LSC and IR regions have been identified. Based on the gene content of the IR region in photosynthetic *Corybas*, the IR analogue for *C. cryptanthus* was assumed to include the region between the *rps3* and *rpl32* genes, inclusive, and the intergenic spacer between *rpl32* and *trnK* genes, which is the first gene within the LSC region in photosynthetic *Corybas*. The IR analogues in the two *C. cryptanthus* accessions are ~4000 nt shorter than the IR regions in photosynthetic *Corybas* plastomes (Table 2.3). This reduction is primarily due to deletions in intergenic regions. The LSC analogue was assumed to include the region between the *trnK* gene and the intergenic spacer up to the start of *rps3*, inclusive. The LSC analogues in the two *C. cryptanthus* accessions are similar, but not identical, in length (Table 2.3). Both are, however, ~30,000-31,000 nt smaller than the LSC regions in photosynthetic *Corybas* plastomes. This reduction is attributed to a combination of extensive gene loss and deletions involving intergenic spacer regions.

2.3.2 GC content

All the sequenced plastomes were broadly similar in terms of overall GC content (Table 2.3). The GC content of the *C. cryptanthus* plastomes (i.e., 36.60%) does, however, differ from that of the remaining *Corybas* plastomes, including *C.* "Waitarere white" (i.e., 38.00-38.20%).

In those plastomes that retain a quadripartite structure, there were clear differences in GC content between the IR, LSC, and SSC regions. Specifically, the GC content was consistently higher for the IR (40.30-43.90%) than either of the

39

single copy regions; moreover, GC content was higher for the LSC region (35.10-36.20%) than the SSC region (27.40-33.20%). The GC content also differs between the IR and LSC analogues in *C. cryptanthus*. Consistent with the trend in photosynthetic Diurideae, the GC content of the IR analogue is higher than that of the LSC analogue (Table 2.3).

2.3.3 Gene content

The plastomes of the photosynthetic Diurideae and *C.* "Waitarere white" all have similar gene content (Table 2.4). Specifically, they have 68-69 protein-coding, four ribosomal RNA, and 21 transfer RNA genes. Full complements of the ribosomal protein (*rps* and *rpl*) genes, ATP synthase (*atp*) genes, plastid-encoded RNA polymerase (*rpo*) genes, and housekeeping genes (i.e., *ycf1*, *ycf2*, maturase K (*matK*), Clp protease proteolytic subunit (*clpP*), acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (*accD*), and translation initiation factor IF-1 (*infA*)) are noted. These plastomes also retain putatively functional copies of all but one of the photosynthesis-related genes. Specifically, full complements of the photosystem I (*psa*), photosystem II (*psb*), and cytochrome b6/f complex (*pet*) gene complexes as well as the chloroplast envelope membrane protein (*cemA*), *ccsA*, and the hypothetical chloroplast RF3 (*ycf3*) and RF4 (*ycf4*) genes. The exception is the hypothetical chloroplast RF15 (*ycf15*) gene that although classed as photosynthesis-related has unknown function and importance. This gene is absent from all photosynthetic *Corybas* plastomes and that of *C.* "Waitarere white". Furthermore, with a few exceptions due to differences in the position of the IR boundaries, gene order and orientation are also maintained across all these plastomes.

In contrast, the *C. cryptanthus* plastomes contain 28 protein-coding, four rRNA, and 18 tRNA genes (Figure 2.2; Table 2.4). Genes disrupted or absent relative to the other sequenced *Corybas* accessions include all six *atp* and four *rpo* genes, most of the photosynthesis-related genes, and the ribosomal protein L33 (*rpl33*) gene. Intact and putatively functional genes include all the housekeeping genes, all but one of the *rpl* and *rps* genes and two photosynthesis-related genes (i.e., photosystem II reaction centre protein M (*psbM*) and cytochrome b6-f complex subunit 5 (*petG*)). The three tRNA genes absent from *C. cryptanthus* are those for alanine (*trnA*), serine (*trnS*), and valine (*trnV*). In most cases, sequences recognisable as belonging to the absent genes were not recovered, even when using a low

Table 2.4 Plastome gene content of *Corybas* and related taxa

	Chiloglottis cornuta	*Microtis unifolia*	photosynthetic *Corybas* spp.	*Corybas* "Waitarere white"	*Corybas cryptanthus*
NADH dehydrogenase genes					
NADH dehydrogenase subunits	0/11 present	0/11 present	0/11 present	0/11 present	0/11 present
Photosynthesis genes					
cytochrome	6/6 present	6/6 present	6/6 present	6/6 present	1/6 present; *petG* present, *petL* pseudogene
photosystem I	5/5 present	5/5 present	5/5 present	5/5 present	0/5 present
photosystem II	15/15 present	15/15 present	15/15 present	15/15 present	1/15 present; *psbM* present, *psbN, -T* pseudogenes, *psbI* gene remnant
other photosynthesis genes	6/6 present	6/6 present	5/6 present; *ycf15* absent	5/6 present; *ycf15* absent	0/6 present
RNA polymerase genes					
RNA polymerase subunits	4/4 present	4/4 present	4/4 present	4/4 present	0/4 present
ATP synthase genes					
ATP synthase subunits	6/6 present	6/6 present	6/6 present	6/6 present	0/6 present
Ribosomal protein genes					
ribosomal protein L subunits	9/9 present	9/9 present	9/9 present	9/9 present	8/9 present; *rpl33* gene remnant
ribosomal protein S subunits	12/12 present	12/12 present	12/12 present	12/12 present	12/12 present
RNA genes					
ribosomal RNA genes	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present
transfer RNA genes	21/21 present	21/21 present	21/21 present	21/21 present	18/21 present; *trnA, trnS, trnV* absent
House-keeping genes					
	6/6 present	6/6 present	6/6 present	6/6 present	6/6 present

similarity threshold (i.e., 35%). In some cases, adjacent genes appear to have been lost along with the intervening sequences suggesting that these losses are the result of a single large deletion. Examples include sections of the plastome spanning cytochrome b559 subunit alpha (*psbE*) to photosystem II reaction centre protein J (*psbJ*), photosystem I reaction centre subunit VIII (*psaI*) to *ycf4*, and *ycf3* to photosystem I P700 chlorophyll a apoprotein A1 (*psaA*). In others, deletions remove part of a gene sequence with the retained portion recognisable as part of the gene. For example, in the gene encoding the cytochrome b6 subunit of the cytochrome b6/f complex (*petB*) the entire first exon and start of the second have been deleted by what appears to be one large deletion, which also removes the intergenic sequence between *petB* and the photosystem II subunit H gene (*psbH*). The remaining portion of the second exon has 77.5% nucleotide identity with the corresponding sequence in *C.* "Remutaka".

No putatively functional *ndh* genes were identified for any of the plastomes characterised in this study (Table 2.4). In *Corybas,* all the *ndh* genes are considered absent; sequences of sufficient length (i.e., >40% the length of presumably functional copies) or amino acid sequence identity (i.e., >60% identity with presumably functional copies) to be considered pseudogenes or gene remnants were not identified. The *M. unifolia* plastome contains a NADH-quinone oxidoreductase subunit A (*ndhC*) pseudogene, but no putatively functional *ndh* genes. In this case, the inferred polypeptide sequence of the pseudogene is ~50% shorter than is typical of *ndhC* (i.e., 67 amino acids in *M. unifolia* versus ~120 amino acids in other orchid species) due to the introduction of a premature stop codon together with several in-frame deletions. Despite this, the nucleotide sequence beyond this stop codon retains high identity to other orchid *ndhC* gene sequences (i.e., ~83% over the predicted length). The *Ch. cornuta* plastome also contains a NADH-quinone oxidoreductase subunit K (*ndhE*) gene remnant. Again, the inferred polypeptide sequence is shorter than a putatively functional version of the gene (i.e., 41 amino acids in *Ch. cornuta* versus ~101 amino acids in other orchid species) due to a 2 nt insertion that generates a premature stop codon. Prior to this stop codon, the inferred polypeptide sequence is highly similar to putatively coding versions of the gene (i.e., ~95% identity) with little nucleotide similarity to *ndhE* gene sequences after it.

2.3.4 Single nucleotide polymorphisms and indels

Pairwise comparisons of photosynthetic *Corybas* suggest that the LSC region is more variable than the IRs (Table 2.5). Considering all pairwise comparisons, there were 2-373 SNPs within IRs (0.02-1.29% of sequence positions within an IR) and 26-2051 for the LSC regions (0.03-2.76% of sequence positions within the LSC region). This overall pattern was also apparent for individual comparisons. For example, there are seven and 26 SNPs in the IR and LSC region, respectively, for the comparison of *C.* "Remutaka" and *C. obscurus*. Results for comparisons involving *C.* "Waitarere white" were consistent with those involving the photosynthetic *Corybas*. In contrast, comparisons of photosynthetic *Corybas* to *C. cryptanthus* suggest a substantially larger number of SNPs. Moreover, the IR and LSC region analogues in *C. cryptanthus* are more similar in terms of SNP number than pairwise comparisons involving photosynthetic *Corybas* (Table 2.5).

Results were similar for indels, with the LSC region having more indels than the IR (Table 2.6). Considering all pairwise comparisons for photosynthetic *Corybas* there were 1-92 indels inferred within IRs and 13-339 for the LSC region. Again, this overall pattern was apparent for the individual comparisons. Results for *C.* "Waitarere white" were consistent with those of the photosynthetic *Corybas*. Larger numbers of indels were identified in pairwise comparisons between the IRs of photosynthetic *Corybas* and the IR analogues of *C. cryptanthus* (i.e., 98-134 indels). When compared to one another, the *C. cryptanthus* LSC analogues have 22 indels and the IR analogues have none.

2.3.5 Gene variability

Putatively functional protein-coding genes vary little in overall length in photosynthetic *Corybas* plastomes (Table 2.7). Most (79%) are the same length in all species and all except one vary by <30 nt. The two longest genes, *ycf1* and *ycf2*, are also the most variable in length, ranging from 5052-5172 and 6438-6465 in *ycf1* and *ycf2*, respectively. Despite this, gene length appears not to be strongly associated with gene length variation in *Corybas*. For example, the plastid-encoded RNA polymerase β subunit (*rpoB*) is 3213 nt long in all photosynthetic *Corybas*, whereas the much shorter ribosomal protein L20 (*rpl20*) varies from 360 to 390 nt in length.

None of the protein-coding genes are identical at the nucleotide sequence level across the sequenced photosynthetic *Corybas* plastomes. However, for most

Table 2.5 Summary of SNPs for comparisons

Pairwise comparisons of *Corybas* with intact quadripartite plastome

Species	IR	IR %	LSC	LSC %
Corybas cheesemanii	335-372	1.16-1.29	1845-2051	2.50-2.62
Corybas diemenicus	167-373	0.58-1.29	1086-2051	1.48-2.77
Corybas dienemus	6-355	0.02-1.22	27-1937	0.04-2.62
Corybas iridescens	6-354	0.02-1.22	27-1936	0.04-2.62
Corybas macranthus	30-347	0.10-1.19	245-1878	0.33-2.55
Corybas obscurus	7-335	0.02-1.16	27-1845	0.03-2.50
Corybas "Remutaka"	7-358	0.02-1.22	27-1846	0.03-2.50
Corybas vitreus	24-339	0.08-1.16	180-1845	0.24-2.50
Corybas "Waitarere white"	6-286	0.02-0.98	70-1878	0.09-2.54

Pairwise comparisons of *C. cryptanthus* with other *Corybas*

Species	IR analogue	IR analogue %	LSC analogue	LSC analogue %
Corybas cryptanthus [b]	670-789	2.72-3.21	–	–
Corybas cryptanthus [a]	688-814	2.80-3.31	–	–

Pairwise comparisons of the two *C. cryptanthus* accessions

Species	IR analogue	IR analogue %	LSC analogue	LSC analogue %
Corybas cryptanthus	12	0.045	18	0.042

[a] This accession was collected in East Harbour Regional Park, Wellington.

[b] This accession was collected in Omoana, Taranaki.

Table 2.6 Summary of indels for comparisons

Pairwise comparisons of *Corybas* with intact quadripartite plastome

Taxa	IR	LSC
Corybas cheesemanii	77-92	322-339
Corybas diemenicus	42-90	200-339
Corybas dienemus	1-91	13-339
Corybas iridescens	1-92	13-330
Corybas macranthus	6-85	73-337
Corybas obscurus	2-77	25-322
Corybas "Remutaka"	2-89	25-326
Corybas vitreus	4-90	71-325
Corybas "Waitarere white"	6-64	31-303

Pairwise comparisons of *C. cryptanthus* with other *Corybas*

Taxa	IR analogue	LSC analogue
Corybas cryptanthus [b]	99-134	–
Corybas cryptanthus [a]	100-139	–

Pairwise comparisons of the two *C. cryptanthus* accessions

Taxa	IR analogue	LSC analogue
Corybas cryptanthus	0	22

[a] This accession was collected in East Harbour Regional Park, Wellington.

[b] This accession was collected in Omoana, Taranaki.

Table 2.7 Statistics for gene variation in *Corybas* species

	Mean and range (in parentheses) of gene length (nt)	Mean and range (in parentheses) of gene length variation (nt)	Mean and range (in parentheses) of mean nucleotide identity (%)	Mean and range (in parentheses) of mean amino acid identity (%)	Mean and range (in parentheses) of pairwise positives (%)
Genes not disrupted in any of the sampled species	816 (96-6465)	8.7 (0-120)	98.9 (94.9-99.9)	98.7 (95.6-100)	99.2 (96.5-100)
Genes disrupted in at least one of the sampled species	855 (90-4122)	1.8 (0-24)	98.9 (92.3-99.8)	99.2 (94.5-100)	99.5 (95.6-100)
Genes located in the IR	1456 (180-6456)	17.4 (0-120)	98.9 (97.1-99.8)	98.7 (95.6-100)	99.1 (98.9-100)
Genes located in the LSC	734 (90-4122)	2.2 (0-45)	98.9 (92.3-99.8)	99.1 (96.6-100)	99.5 (99.3-100)
Genes in the photosynthesis-gene class	589 (90-2253)	1.8 (0-24)	99.0 (92.3-99.8)	99.2 (94.5-100)	99.5 (95.6-100)
Genes in the non-photosynthesis gene class	1024 (114-6465)	7.1 (0-120)	98.8 (96.0-99.9)	98.8 (95.6-100)	99.3 (96.5-100)

(>75%) sequence identity is above 97% at both nucleotide and amino acid sequences. Across the sequenced photosynthetic *Corybas* plastomes, only seven genes have average pairwise nucleotide identities that fall below 96%; these were ATP synthase subunit a (*atpI*), ribosomal protein S11 (*rps11*), ribosomal protein S15 (*rps15*), ribosomal protein S16 (*rps16*), *rpl20*, cytochrome b6/f complex subunit 6 (*petL*) and *ycf1*. The *atpI, petL,* and *rps16* genes were the least conserved, with average pairwise nucleotide identities ~92%. Despite this, these genes often retain identity at the amino acid level. For example, *atpI* has an amino acid identity of 99.8%, implying that most nucleotide substitutions are silent.

In the *C. cryptanthus* plastomes ~60% of genes that remain putatively functional were the same length as in photosynthetic congeners. Twelve genes were shorter and one was longer. Eight of these genes differed by less than 35 nt; two, *ycf1* and *ycf2*, were 288-408 and 222-249 nt shorter, respectively, while *matK* was 57 nt longer. In all but one case, gene lengths were identical in both *C. cryptanthus* plastomes; the exception is *infA*, which differed by 9 nt, although this gene did not vary in length between other *Corybas* plastomes.

Comparisons of photosynthetic *Corybas* gene sequences found no significant difference in terms of sequence length and similarity at the nucleotide or amino acid levels between genes functionally lost from the *C. cryptanthus* plastome and those retained in *C. cryptanthus* (two sample t-tests, t = -1.51-1.60, *p*-value = 0.116-0.793, DF = 30-65). These results suggest that the underlying variability of a gene does not predispose it to disruption in *C. cryptanthus*. Results were similar for gene classes. In this case, two groups were considered, the photosynthesis gene class (e.g., *psa, psb, cemA*) and a group containing all the remaining genes – this included only gene classes lost after photosynthesis genes in the Barrett and Davis model (e.g., *rpo, atp*) since the *ndh* genes are lost from all sequenced plastomes. Differences in length variation and residue identity were not statistically significant (two sample t-tests, t = -0.94-1.55, *p*-value = 0.128-0.477, DF = 44-61). For the photosynthetic *Corybas* plastomes examined, nucleotide and amino acid identities do not differ significantly between the IR and LSC (two sample t-tests, t = -1.12-0.03, *p*-value = 0.290-0.975, DF = 10-14). Given the variation in gene content for the SSC it was not possible to include this in comparisons.

2.3.6 Repeats

Short repeat sequences, 15-29 nt in length, are more numerous in the plastomes of Diurideae than repeats 30 nt or longer (Tables 2.7 and 2.8). Specifically, in pairwise comparisons of photosynthetic Diurideae plastomes there are 92-157 repeat sequences that are 15-29 nt in length and 0-3 of 30 nt or longer. Repeat numbers for the *C.* "Waitarere white" plastome are consistent with those of the photosynthetic *Corybas*. In contrast, the plastomes of *C. cryptanthus* have fewer repeat sequences 15-29 nt in length (i.e., 54-58 c.f. 92-108). However, while *C. cryptanthus* plastomes contained fewer small repeats than plastomes of photosynthetic Diurideae, the plastomes are also smaller in size, such that the contribution of the repeats to genome size is consistent with that of other *Corybas* species. The plastomes of *C. cryptanthus* contain more repeats 30 nt or longer than those of photosynthetic Diurideae (i.e., 4 c.f. 0-3 repeats). Moreover, the *C. cryptanthus* repeats are 34-85 nt in length, whereas they are 30-40 nt in length for the photosynthetic Diurideae.

In the photosynthetic plastomes, all repeats 30 nt or longer are located within the LSC and, with one exception in *Ch. cornuta*, all occur within intergenic spacers (Table 2.9). Most of these repeats are palindromes. Some are shared by multiple accessions.

Table 2.8 Frequency of repeats 15-29 nt in the Diurideae plastome sequences

Species	Number of repeats	Percentage plastome occupied by repeats
Chiloglottis cornuta	157	0.12%
Corybas cheesemanii	95	0.07%
Corybas cryptanthus [b]	54	0.08%
Corybas cryptanthus [a]	58	0.08%
Corybas diemenicus	108	0.08%
Corybas dienemus	98	0.07%
Corybas iridescens	100	0.07%
Corybas macranthus	101	0.08%
Corybas obscurus	92	0.07%
Corybas "Remutaka"	95	0.07%
Corybas "Waitarere white"	102	0.08%
Corybas vitreus	101	0.07%
Microtis unifolia	97	0.08%

[a] This accession was collected in East Harbour Regional Park, Wellington

[b] This accession was collected in Omoana, Taranaki

Table 2.9 Size and location of repeats ≥ 30 nt in the Diurideae plastome sequences

Species	Number of repeats	Repeat sizes and, in parentheses, size range	Location of repeat					
			Functional class			Genome region		
			Intron	Intergenic spacer	Coding sequence	LSC	SSC	IR
Chiloglottis cornuta	2	32-34 (2)	1	1	0	2	0	0
Corybas cheesemanii	0	–	–	–	–	–	–	–
Corybas cryptanthus[a]	4	34-64 (30)	1[c]	3[c]	1	3[d]	-	2[d]
Corybas cryptanthus[b]	4	34-85 (51)	1[c]	3[c]	1	3[d]	-	2[d]
Corybas diemenicus	1	36	0	1	0	1	0	0
Corybas dienemus	2	30-36 (6)	0	2	0	2	0	0
Corybas iridescens	2	30-36 (6)	0	2	0	2	0	0
Corybas macranthus	2	30-32 (2)	0	2	0	2	0	0
Corybas obscurus	2	32-33 (1)	0	2	0	2	0	0
Corybas "Remutaka"	2	32-33 (1)	0	2	0	2	0	0
Corybas "Waitarere white"	1	33	0	1	0	1	0	0
Corybas vitreus	0	–	–	–	–	–	–	–
Microtis unifolia	3	34-40 (6)	0	3	0	3	0	0

[a] This accession was collected in East Harbour Regional Park, Wellington

[b] This accession was collected in Omoana, Taranaki

[c] One copy of one or more repeats is located within an intron and the other in an intergenic spacer.

[d] One copy of one repeat is located within the LSC region and the other in the IR.

For example, *C. obscurus*, *C.* "Remutaka", *C. vitreus* and *C. macranthus* all share a 33 nt inverted repeat located within the intergenic spacer between the *rpl22* and *psbA* genes. Three of the four *C. cryptanthus* repeats in the 30 nt or longer size class are located within the LSC analogue; the fourth is shared between the LSC region and IR analogues. Two of these large repeats are in intergenic spacers. One has one repeat copy within an intergenic spacer and the other within the *ycf1* coding region. The fourth repeat is located within the intron of ribosomal protein L16 (*rpl16*).

2.4 Discussion

To understand the evolutionary trajectory of plastomes in mycoheterotrophic plant species it is useful to compare them to the plastomes of photosynthetic relatives. Comparative analyses of plastomes from 10 photosynthetic and three non-photosynthetic members of the Diurideae – the majority from the genus *Corybas* – are discussed, with a focus on what these analyses suggest about the evolution of plastomes in mycoheterotrophic plants.

2.4.1 Plastomes of the photosynthetic Diurideae
The plastomes of the photosynthetic Diurideae investigated here are broadly similar. For example, all have the quadripartite structure typical of angiosperms (Wicke et al., 2011) (Figure 2.3; Appendix II), similar GC contents (Table 2.3), and similar numbers of repeats (Tables 2.8; 2.9). These plastomes also have highly similar gene complements (Table 2.4), with most genes conserved in terms of length and residue identity, especially among representatives of *Corybas*.

Despite the overall similarity of the plastomes, the SSC region varies dramatically, both in terms of size and gene content, between taxa (Table 2.3; Figure 2.3; Appendix III). The SSC regions of *Ch. cornuta* and *M. unifolia* are 8719 and 10,856 nt in length, respectively; both contain complete coding sequences for five or six genes. In contrast, the SSC regions of photosynthetic *Corybas* plastomes are 1049-1319 nt in length and contained the complete coding sequences of, at most, two genes. Reduction of the SSC region in these taxa is associated with expansion of the IR and, therefore, of the plastome overall (Table 2.3). The SSC regions of the photosynthetic *Corybas* are the smallest reported for a photosynthetic plant to date.

That of *Lamprocapnos spectabilis* (L.) Fukuhana (Papaveraceae) is the next smallest, at 1,645 nt (Cho et al., 2018; Seongjun Park et al., 2018).

The plastomes of the photosynthetic Diurideae retain, with one exception, a common gene set (Table 2.4). Both *Ch. cornuta* and *M. unifolia* have a complement of 94 protein-coding, tRNA and rRNA genes; all the *ndh* genes have been functionally lost from these plastomes. The photosynthetic *Corybas* have lost an additional gene, *ycf15*. In *Corybas* protein-coding genes exhibit contrasting patterns of variability. Some are highly similar across species in terms of both length and sequence identity whereas others vary markedly. Differences in the extent to which length and sequence identity varied were not statistically significant when considered in terms of the Barrett and Davis (2012) gene classes. That is, genes in the photosynthesis gene class, which the Barrett and Davis (2012) model suggest should be lost following the *ndh* genes, were not significantly more variable than genes expected to be lost in the later stages of the model. This result implies that complete loss of the *ndh* genes has not significantly reduced selection pressure for the retention of the photosynthesis genes. Loss of the *ndh* genes does not appear to have committed these plastomes to further genome degradation. This is consistent with the observation that numerous plant species remain photosynthetic despite the functional loss of the *ndh* genes from the plastome (Blazier et al., 2016; Kim & Chase, 2017; Silva et al., 2018).

Pairwise comparisons of SNP diversity for the photosynthetic species are consistent with published *Corybas* phylogenies (e.g., Lehnebach et al., 2016; Lyon, 2015). For example, Lehnebach et al. (2016) reported that *C.* "Remutaka" and *C. obscurus* were sister, and there are fewer SNPs (i.e., 33 SNPS) between these two accessions than in comparisons of these accessions to any other photosynthetic *Corybas* (i.e., 204-2204 SNPs). Similarly, *C. diemenicus* and *C. cheesemanii*, which are progressively more distantly related to *C.* "Remutaka" and *C. obscurus* in the Lyon (2015) phylogeny, have progressively larger numbers of SNPs; 1263-1264 SNPs for *C. diemenicus* and 2180-2204 for *C. cheesemanii*. Indel number exhibits a similar pattern.

2.4.2 The plastomes of non-photosynthetic Corybas

The plastomes of the two non-photosynthetic *Corybas* taxa sampled differ markedly. That of the *C.* "Waitarere white" accession was highly similar to those of

photosynthetic relatives whereas, although the plastomes from the two
C. cryptanthus accessions were alike, they differed from photosynthetic relatives.

The plastome of *C. cryptanthus* is approximately half the size of the
remaining *Corybas* plastomes (Table 2.3). The loss of one IR copy accounts for
about half of the nearly 65,000 nt lost; a single IR analogue is identifiable. This IR
analogue contains all but one of the genes common to the IR of photosynthetic
Corybas. These genes are also in the same order, orientation, and overall position
within the plastome. Much of the remaining size reduction, ~28,000 nt, in
C. cryptanthus is due to deletions from the LSC analogue. Deletions from intergenic
spacers within the IR analogue and of the entire SSC region account for ~3400 nt and
~1100 nt, respectively. The loss of one IR copy and the SSC region also makes the
C. cryptanthus plastome structurally different to those of other *Corybas* and may
explain lower GC content estimates for *C. cryptanthus*. In photosynthetic *Corybas*,
the GC content of the IRs was higher than that of the single copy regions and the loss
of one IR copy would have a substantial impact on overall GC content in
C. cryptanthus (Table 2.3). Although the *C. cryptanthus* plastome is reduced, the
order of the remaining genes is conserved, suggesting structural rearrangements have
not played a role in plastome change.

Gene content in the *C. cryptanthus* plastome corresponds well with that of
similarly sized plastomes from other heterotrophic plants (Appendix III). Most
plastomes of comparable size are from holoparasitic Orobanchaceae and most retain
the typical quadripartite structure. Except for *Orobanche latisquama* (F.W.Schultz)
Batt (Wicke et al., 2013), which retains a single *ndh* gene, functional *ndh* and *rpo*
genes are almost entirely absent from these plastomes, with pseudogenes and gene
remnants also uncommon. As in *C. cryptanthus*, these plastomes also retain few
photosynthesis-related genes. For the remaining gene classes of Barrett and Davis
(2012), gene retention is often higher. The number of functional ribosomal protein
and house-keeping genes retained is relatively consistent; one to six ribosomal
protein and up to two housekeeping genes have been lost from these plastomes.
However, the numbers of functional *atp* and tRNA genes are more variable. For
example, the full set of *atp* genes is retained in *Orobanche ramosa* (L.) Pomel.
(Wicke et al., 2013) but these are entirely lost, at least functionally, in
C. cryptanthus.

Beyond these similarities, the plastome of *C. cryptanthus* also appears to broadly fit the Barrett and Davis (2012) model of gene loss for non-photosynthetic plant plastomes. The *C. cryptanthus* plastome lacks all *ndh*, *rpo*, and *atp* genes as well as most of the photosynthesis genes, but retains most tRNA and ribosomal protein genes as well as all housekeeping genes (Table 2.4). This level of gene loss suggests that *C. cryptanthus* is in the final stage of the Barrett and Davis (2012) model. However, that two putatively functional photosynthesis genes (i.e., *psbM* and *petG*) are retained is somewhat at odds with the predictions of the model. Photosynthesis genes are lost in the second step of the model, whereas loss of *rpo* and *atp* genes occurs at steps three and four, respectively. No *rpo* or *atp* genes or pseudogenes remain in the *C. cryptanthus* plastome, which implies substantial overlap between the steps. That is, not all genes of a given class need be lost before gene losses involving the subsequent class can begin.

One possibility is that the observed overlap in the relative timing of gene loss is linked to the size of the photosynthesis gene class. This class contains 32 genes, of which 26 are subunits of three protein complexes (i.e., cytochrome, photosystem I and II). The entire photosynthesis gene class may not, therefore, need to be lost before selection on later gene classes is sufficiently relaxed for disruption to begin. For example, loss of only one or two subunit genes may be needed before functionality of the complex, and any pressure it exerts for the retention of later gene classes, is lost.

In contrast to *C. cryptanthus*, there is no indication that the loss of chlorophyll and, presumably, photosynthesis is associated with plastome change in *C.* "Waitarere white". The structure, gene content and size of this plastome are consistent with those from photosynthetic *Corybas*. This plastome is very similar to representatives of the Trilobus aggregate, in this study represented by *C. obscurus* and *C.* "Remutaka". Previous phylogenetic studies (Lehnebach et al., 2016) have suggested that *C. obscurus* and *C.* "Remutaka" are very closely related to one another, a suggestion supported by the plastome sequencing conducted in this study. Both these taxa and a third member of the Trilobus aggregate, currently referred to as *C.* "pygmy", occur in the Waitarere forest, along with several other *Corybas* species (i.e., *C. vitreus* and *C. macranthus*). The chloroplast genome of the albino individual is similar at the nucleotide sequence level, but not identical, to those of the *C. obscurus* and *C.* "Remutaka" accessions characterised in this study. One possible

explanation for *C.* "Waitarere white" is that it is an albino form of a Trilobus aggregate species found in the Waitarere forest. If so, the lack of gene loss from the *C.* "Waitarere white" plastome suggests these plants have recent origins. The retention of leaves in these specimens also implies that the loss may be recent. That is, there has not been enough time for these plants to acquire the morphology typical of holomycoheterotrophs. That said, the complete loss of pigmentation in these plants implies a qualitatively different situation to that in *C. cryptanthus*; one that may not lead to the persistence of an evolutionary distinct lineage.

Sequencing of further *C.* "Waitarere white" accessions and representatives of the photosynthetic taxa from the Waitarere forest is now needed. Such analyses would allow the origins of *C.* "Waitarere white" to be evaluated in more detail; for example, whether a local population of photosynthetic *Corybas* has been the source of albino individuals and whether albino individuals persist over time or arise *de novo*. Since the plastome is unchanged in the accession sequenced here, an evaluation of genetic variation at nuclear loci associated with pigmentation and photosynthesis may provide mechanistic insights. Further analyses could also incorporate variegated specimens, which may be caused by similar phenomena. Variegation has not been widely studied in Orchidaceae, but in *Arabidopsis thaliana* L., both variegation and albinism can be caused by disruptions to cytonuclear interactions. For example, disruptions to nuclear genes encoding subunits that associate with the plastid-encoded RNA polymerase are associated with both albinism (Yu et al., 2014) and variegation (Hu et al., 2015).

2.4.3 Implications of intraspecific variation between C. cryptanthus plastomes
The plastomes of the two *C. cryptanthus* accessions are similar in both size and structure. Specifically, both have lost a single copy of the IR and differ in length by 76 nt. Both also have the same set of 51 putatively functional genes (i.e., 29 protein-encoding, 18 tRNA, and 4 rRNA genes), 3-4 gene remnants (i.e., ATP synthase epsilon subunit (*atpE*), photosystem II reaction centre protein I (*psbI*), photosystem II reaction centre protein Z (*psbZ*), and *rpl33* in the Taranaki accession), and one pseudogene (i.e., photosystem II reaction centre protein T (*psbT*)). For all but two putatively functional protein-encoding genes, the nucleotide sequences are identical in the two accessions; as are all but one of the gene remnants (i.e., *rpl33*). The two genes that differ at the nucleotide level between the *C. cryptanthus* accessions are

infA and *matK*. In *matK*, a single nucleotide polymorphism leads to a single amino acid substitution, whereas in *infA*, the insertion of a single nucleotide causes a frameshift that abolishes the usual stop codon and adds three amino acids to the protein product before an alternative stop codon terminates the gene.

The most obvious difference between the two *C. cryptanthus* plastomes involves size; they differ in length by 76 nt. This is similar to size differences between recognised species of photosynthetic *Corybas* (e.g., *C. iridescens* and *C. dienemus*, which differ by 74 nt). The size difference between *C. cryptanthus* accession is due to 22 indels in the LSC region analogue. By comparison, pairwise alignments of the remaining Nematoceras clade taxa identified 2-20 indels in the IRs and 13-116 in the LSC region. In addition, the *C. cryptanthus* plastomes differ by 30 SNPs, 12 in the IR analogue and 18 in the LSC region analogue. This is again within the range of variation for pairwise comparisons between the other members of the Nematoceras clade; 6-72 SNPs in the IRs and 26-542 in the LSC region.

That both *C. cryptanthus* share the loss of one IR copy as well as the same gene losses, yet have diverged at the sequence-level, suggests that these losses occurred prior to the establishment of these populations. It would seem unlikely for such a complex series of losses to have occurred in parallel in geographically isolated populations. At face value, the extent of sequence difference between the *C. cryptanthus* plastomes suggests that divergence in this species has occurred over a similar timeframe to diversification of the Nematoceras clade. Molecular age estimates from an analysis of nrITS sequences imply that the Nematoceras clade began to diversify ~4 million years ago (Lyon, 2015). That said, a wider comparison of SNP variation suggests differences between the *C. cryptanthus* plastomes may have arisen more recently. Molecular age estimates from Lyon (2015) suggest that *C. cryptanthus* last shared a common ancestor with the Nematoceras clade ~9 million years ago and that, together with the Corysanthes clade, they shared a common ancestor with the large clade to which *C. cheesemanii* belongs ~15 million years ago. In contrast, in pairwise comparisons of the IRs from *C. cheesemanii* and members of the Nematoceras clade there were 167-358 SNPs but 670-789 between the IRs of the Nematoceras clade taxa and IR analogue of *C. cryptanthus*. This implies an increased nucleotide substitution rate, at least within the IR analogue of *C. cryptanthus*, perhaps reflecting loss of the second IR copy from this genome. Accelerated rates of sequence evolution have been reported for other heterotrophic plant species (Wicke

et al., 2016). This suggests that despite having been present in New Zealand for ~9 million years and potentially having been non-photosynthetic for an extended time period, the two *C. cryptanthus* populations sampled in the present study diverged only relatively recently.

The extent to which the two *C. cryptanthus* populations are genetically isolated is difficult to assess. The cryptic nature of mycoheterotrophs such as *C. cryptanthus* makes assessing whether there are morphological, physiological, and habitat differences between populations more challenging. These populations are separated by several hundred kilometres, which combined with the short stature of plants and preference for forested habits makes ongoing gene flow between them unlikely. Moreover, while long distance dispersal is now widely accepted as having an important role in the establishment of many plant groups in New Zealand (e.g., Winkworth et al., 2002), the effectiveness of this mechanism for maintaining gene flow over longer time scales has not been rigorously tested.

Plastomes from multiple individuals of a mycoheterotrophic species have been compared in only a few investigations, most focusing on Ericaceae. Specifically, Ravin et al. (2016) found that the plastomes of two *M. hypopitys* L. (Ericaceae) accessions from geographically distinct locations differed in size, gene content, and structure. Less dramatic differences were also found by Schelkunov et al. (2015) between the plastomes of *Epipogium roseum* (D. Don). Lindl (Orchidaceae) from three geographically distinct locations, with similar findings reported for three further Ericaceae species (Braukmann et al., 2017). Cryptic speciation has been suggested as a possible mechanism for these differences (Braukmann et al., 2017; Schelkunov et al., 2015).

Despite differing from each other by a similar number of SNPs and indels as are observed between species of photosynthetic *Corybas*, the two *C. cryptanthus* plastomes differ less in terms of length, nucleotide identity and indels than between different accessions of other mycoheterotrophic species. Whether or not the populations represented by these two *C. cryptanthus* accessions should be considered different species remains to be tested. Nevertheless, the present analyses do suggest that they have diverged from one another, perhaps in isolation, following the transition to mycoheterotrophy.

2.4.4 Impact of IR boundary shifts

In the Diurideae investigated, increases in the overall size of the plastome are driven by expansion of the IRs at the expense of the SSC region. For example, 72% of the size difference between *Ch. cornuta* and *C. dienemus* is explained by the expansion of the IR. In some orchid species, reduction of the SSC has been attributed to *ndh* gene loss, as many of these genes are clustered within the SSC region (Kim et al., 2018). However, in this case, reduction of the SSC region is not the result of gene loss. The plastomes of *M. unifolia* and *Ch. cornuta* retain almost the same gene set as that of photosynthetic *Corybas*; all having lost the entire suite of *ndh* genes. The one gene content difference between photosynthetic *Corybas* and these two related taxa is loss of the *ycf15* gene, which is in the LSC region and so would not contribute to reduced SSC region size. Instead, most of the size differences between photosynthetic *Corybas* and their relatives are explained by changes in the boundaries between the SSC and IR regions. Typical of other angiosperms, in *Ch. cornuta* and *M. unifolia,* the IR$_A$ boundary sits within the *ycf1* gene and just outside it (i.e., the gene is fully within the SSC region), respectively. In contrast, this boundary has shifted so that the entire *ycf1* gene as well as neighbouring genes typically located within the SSC region (i.e., *rps15* and *psaC*) fall within the IR in photosynthetic *Corybas*.

The expansion of the IR into the SSC region dramatically reduces the size of the SSC region without decreasing overall gene content. Expansion of the IR is also responsible for dramatic reductions in SSC length in the plastomes of the photosynthetic *Lamprocapnos spectabilis* (Seongjun Park et al., 2018) and hemiparasitic *Pedicularis ishiodoyana* Koids. & Ohwi (Orobanchaceae) (Cho et al., 2018). Loss of the *ndh* genes is suggested to have led to the destabilisation of the boundary between the IR and the SSC region in some photosynthetic plastomes. The physical, rather than functional, loss of these genes, particularly of NADH dehydrogenase subunit 5 (*ndhF*), seems to be important (Kim et al., 2015; Niu et al., 2017). The plastomes of Diurideae appear to fit with this pattern. In *Corybas*, where the IR-SSC boundaries are unstable, the *ndh* genes, including *ndhF,* have been physically lost. In contrast, *M. unifolia* and *Ch. cornuta* each retain a single *ndh* pseudogene and their IR-SSC are typical of Orchidaceae.

The SSC region gene order is the same in *Ch. cornuta* and *M. unifolia*. Comparisons with gene order in the photosynthetic *Corybas* suggest possible

scenarios for IR expansion. A single expansion into the SSC region seems unlikely because in the related Diurideae the *ycf1*, *rps15* and *psaC* genes are located at the opposite end of the SSC region to the *rpl32* gene. For *psaC* and *rpl32* to be adjacent on photosynthetic *Corybas* plastomes requires at least two rounds of expansion. One possibility is that IR expansion has occurred from both ends of the SSC region, albeit at different times. In this case invasion of IR_A into the SSC would first capture the *ycf1*, *rps15* and *psaC* genes; this could have involved a single large expansion or several smaller ones. Expansion at the IR_A boundary would need to be followed by an expansion of IR_B in order to capture the *rpl32* gene. An alternative is that expansion occurred from only one of the boundaries. There is evidence to suggest that within individual plants two plastome isoforms persist; these forms differ in the orientation of the SSC region relative to the remainder of the plastome and are maintained by so-called "flip-flop recombination" (Palmer, 1983; Walker et al., 2015). If so, depending on the orientation of the SSC region, initial expansion of either the IR_A or IR_B could have captured the *ycf1*, *rps15* and *psaC* genes. The subsequent expansion could have then occurred from the same boundary if the orientation of the SSC was that of the alternative isoform.

In contrast, the boundaries between the IR and LSC region are very stable in terms of position within the genome (Figure 2.4). A difference in the stability of the SSC and LSC region boundaries with the IRs is considered characteristic of Orchidaceae (Kim et al., 2015). Consistent with this, gene order and nucleotide sequence similarity indicate that the boundary between the LSC and IR analogues in *C. cryptanthus* is typical of the boundary between IR_A and the LSC region in photosynthetic *Corybas*. Although the exact position of this boundary cannot be found in the absence of IR_B, given the extent of conservation across Diurideae sampled, it seems probable that it would have been between *rps3* and *rpl22* prior to loss of one IR copy.

Whether IR expansion impacts on the overall stability of the plastome is unknown. Changes in IR length appear to be a general feature of *Corybas* plastome evolution and it is possible that this process also played a role in the IR loss in *C. cryptanthus*. One possibility is that further reduction of the SSC region in ancestors of *C. cryptanthus* resulted in loss of one IR copy. While this idea has yet to be rigorously tested, it is consistent with the observation that only one plastome with a SSC region <1000 nt in length (i.e., *Pedicularis ishiodoyana*) has been reported.

Similar, although not identical, patterns have also been reported. For example, *Aphyllorchis montana* Blume has lost one IR copy but appears to retain genes from the SSC (Feng et al., 2016) whereas *Epipogium aphyllum* Sw. and *Cynomorium coccineum* L. retain two copies of the IR but lack a SSC region (Bellot et al., 2016; Schelkunov et al., 2015). In the absence of plastomes from closely related taxa it is difficult to tell whether IR boundary expansion may have played a role in the evolution of plastome structure in these species; genes typically located in the SSC region may have been incorporated into the IR prior to loss of one copy. This would leave genes that appear to form a SSC analogue, but may have been part of the IR.

It is interesting to note that the final protein-coding gene retained in the SSC of photosynthetic *Corybas*, *ccsA*, appears to be one of the photosynthesis genes which is frequently lost early in plastome degradation. This is evidenced by pseudogenization or loss of *ccsA* from several holomycoheterotrophic plastomes thought to be in the early stages of plastome degradation, such as those of *Corallorhiza* Gagnebin spp. (Barrett et al., 2014) and *Exochaenium oliganthum* (Gilg) Kissling (Darby, 2015). Further work will be needed to determine whether large-scale structural changes, such as those in *C. cryptanthus*, are a consequence of the functional loss of genes or are themselves a key driver of gene loss.

2.5 References

Abadie, J. C., Püttsepp, Ü., Gebauer, G., Faccio, A., Bonfante, P., & Selosse, M. A. (2006). *Cephalanthera longifolia* (Neottieae, Orchidaceae) is mixotrophic: a comparative study between green and nonphotosynthetic individuals. *Canadian Journal of Botany, 84*, 1462-1477.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403-410.

Barrett, C. F., & Davis, J. I. (2012). The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany, 99*, 1513-1523.

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., & Santos, C. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution, 31*, 3095-3112.

Bellot, S., Cusimano, N., Luo, S., Sun, G., Zarre, S., Gröger, A., Temsch, E., & Renner, S. S. (2016). Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasite family Cynomoriaceae in the Saxifragales. *Genome Biology and Evolution, 8*, 2214-2230.

Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M. L., & Ruhlman, T. A. (2016). Variable presence of the inverted repeat and plastome stability in *Erodium*. *Annals of Botany, 117*, 1209-1220.

Braukmann, T. W. A., Broe, M. B., Stefanovic, S., & Freudenstein, J. V. (2017). On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytologist, 216*, 254-266.

Breitwieser, I., Brownsey, P. J., Heenan, P. B., Nelson, W. A., & Wilton, A. D. (2019). Flora of New Zealand Online - Taxon Profile - *Corybas* (based on Flora Committee). Retrieved 9 January 2019 from http://www.nzflora.info/factsheet/taxon/Corybas.html.

Campbell, E. O. (1972). The morphology of the fungal association of *Corybas cryptanthus*. *Journal of the Royal Society of New Zealand, 2*, 43-47.

Cho, W. B., Choi, B. H., Kim, J. H., Lee, D. H., & Lee, J. H. (2018). Complete plastome sequencing reveals an extremely diminished SSC region in hemiparasitic *Pedicularis ishiodoyana* (Orobanchaceae). *Annales Botanici Fennici, 55*, 171-183.

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics, 11*, 485.

Darby, H. (2015). *Plastid genome evolution in partially and fully mycoheterotrophic eudicots*. (Text), Retrieved from https://open.library.ubc.ca/collections/24/items/1.0223133.

de Lange, P. J. (2005). *Corybas cryptanthus*. Retrieved 23 January 2019 from http://nzpcn.org.nz/flora_details.aspx?ID=1391.

Dransfield, J., Comber, J. B., & Smith, G. (1986). A synopsis of *Corybas* (Orchidaceae) in West Malesia and Asia. *Kew Bulletin, 41*, 575-613.

Feng, Y. L., Wicke, S., Li, J. W., Han, Y., Lin, C. S., Li, D. Z., Zhou, T. T., Huang, W. C., Huang, L. Q., & Jin, X. H. (2016). Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biology and Evolution, 8*, 2164-2175.

Gustafsson, A. L. S., Verola, C. F., & Antonelli, A. (2010). Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus Hoffmannseggella(Orchidaceae: Epidendroideae). *BMC Evolutionary Biology, 10*, 177.

Hatch, E. D. (1956). A new species of *Corybas* Salisbury, and a note on some name changes in *Wahlenbergia* Schrader. *Transactions of the Royal Society of New Zealand, Botany 83*, 366.

Hu, F., Zhu, Y., Wu, W., Xie, Y., & Huang, J. (2015). Leaf variegation of Thylakoid Formation1 is suppressed by mutations of specific σ-factors in *Arabidopsis*. *Plant Physiology, 168*, 1066-1075.

Irwin, J. B. (1954). *Corybas saprophyticus*. *Wellington Botanical Society Bulletin, 27*, 2.

Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., & Selosse, M. A. (2005). Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytologist, 166*, 639-653.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*, 1647-1649.

Kim, H. T., & Chase, M. W. (2017). Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). *PloS ONE, 12*, e0187318-e0187318.

Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., & Kim, J.-H. (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PloS ONE, 10*, e0142215.

Kim, H. T., Shin, C. H., Sun, H., & Kim, J. H. (2018). Sequencing of the plastome in the leafless green mycoheterotroph *Cymbidium macrorhizon* helps us to understand an early stage of fully mycoheterotrophic plastome structure. *Plant Systematics and Evolution, 304*, 245-258.

Lehnebach, C. A., Zeller, A. J., Frericks, J., & Ritchie, P. (2016). Five new species of *Corybas* (Diurideae, Orchidaceae) endemic to New Zealand and phylogeny of the Nematoceras clade. *Phytotaxa, 270*, 1-24.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*, 1754-1760.

Lohse, M., Drechsel, O., Kahlau, S., & Bock, R. (2013). OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research, 41*, 575-581.

Lyon, P. S. (2015). *Molecular systematic, biogeography, and mycorrhizal associations in the Acianthinae (Orchidaceae), with a focus on the genus Corybas*. (PhD Thesis), University of Wisconsin-Madison, USA.

Merckx, V. (2012). *Mycoheterotrophy: The Biology of Plants Living on Fungi*. New York: Springer-Verlag.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., & Marshall, D. (2010). Tablet—next generation sequence assembly visualization. *Bioinformatics, 26*, 401-402.

Moore, L. B., & Edgar, E. (1970). *Indigenous Tracheophyta* (Vol. II). Wellington, New Zealand: Government Printer.

Niu, Z., Pan, J., Zhu, S., Li, L., Xue, Q., Liu, W., & Ding, X. (2017). Comparative analysis of the complete plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids.*, 8*, 1713.

Palmer, J. D. (1983). Chloroplast DNA exists in two orientations. *Nature, 301*, 92-93.

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics, 28*, 1420-1428.

Ravin, N. V., Gruzdev, E. V., Beletsky, A. V., Mazur, A. M., Prokhortchouk, E. B., Filyushin, M. A., Kochieva, E. Z., Kadnikov, V. V., Mardanov, A. V., & Skryabin, K. G. (2016). The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC plant biology, 16*, 153–161.

Salmia, A. (1989). General morphology and anatomy of chlorophyll-free and green forms of *Epipactis helleborine* (Orchidaceae). *Annales Botanici Fennici, 26*, 95-105.

Sanmartín, I., Wanntorp, L., & Winkworth, R. C. (2007). West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. *Journal of Biogeography, 34*, 398-416.

Schelkunov, M. I., Shtratnikova, V. Y., Nuraliev, M. S., Selosse, M. A., Penin, A. A., & Logacheva, M. D. (2015). Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution, 7*, 1179-1191.

Selosse, M. A., Faccio, A., Scappaticci, G., & Bonfante, P. (2004). Chlorophyllous and achlorophyllous specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) are associated with ectomycorrhizal septomycetes, including truffles. *Microbial Ecology, 47*, 416-426.

Seongjun Park, Boram, A., & Park, S. (2018). Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion and intraspecific variation. *Scientific Reports, 8*, 13568.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology, 7*, 539-539.

Silva, S. R., Michael, T. P., Meer, E. J., Pinheiro, D. G., Varani, A. M., & Miranda, V. F. O. (2018). Comparative genomic analysis of *Genlisea* (corkscrew plants—Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes. *PloS ONE, 13*, e0190321.

Snedecor, G. W., & Cochrane, W. G. (1989). *Statistical Methods* (Eighth ed.). USA: Iowa State University Press.

Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany, 99*, 349-364.

Twyford, A. D., & Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular Ecology Resources, 17*, 858-868.

Walker, J. F., Jansen, R. K., Zanis, M. J., & Emery, N. C. (2015). Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. *American Journal of Botany, 102*, 1751-1752.

Watkins, R. L. S. (2012). *The biogeography, ecology and endophyte mycorrhiza of the New Zealand Corybas alliance (Orchidaceae) specifically: Nematoceras iridescens (Irwin et Molloy) Molloy, D.L. Jones & M.A. Clem (Species).* (Doctor of Philosophy in Plant Biology), Massey University, Palmerston North, New Zealand.

Wicke, S., Muller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., Renner, S. S., & Schneeweiss, G. M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell, 25*, 3711-3725.

Wicke, S., Müller, K. F., DePamphilis, C. W., Quandt, D., Bellot, S., & Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 9045-9050.

Wicke, S., Schneeweiss, G. M., de Pamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology, 76*, 273-297.

Winkworth, R. C., Wagstaff, S. J., Glenny, D., & Lockhart, P. J. (2002). Plant dispersal N.E.W.S from New Zealand. *Trends in Ecology & Evolution, 17*, 514-520.

Yu, Q. B., Huang, C., & Yang, Z. N. (2014). Nuclear-encoded factors associated with the chloroplast transcription machinery of higher plants. *5*, 316.

Chapter 3

The complete chloroplast genome of the mycoheterotrophic orchid, *Danhatchia australis*: implications for taxonomy and plastome evolution

3.1 Introduction

Danhatchia australis (Hatch) Garay & Christenson (Cranichideae, Goodyerinae) is a terrestrial, mycoheterotrophic orchid that until recently was considered endemic to New Zealand. First described by Hatch (1963), plants now attributed to *Danhatchia* Garay & Christenson were originally described as *Yoania australis* Hatch. *Yoania* Maxim. (Epidendroideae, Calypsoinae) is a small genus that currently comprises four fully mycoheterotrophic species native to Asia (Merckx, 2012). Due to differences in the flower structure (e.g., shorter pedicels, fleshy and obtuse column-wings, and different pollinia numbers) Hatch (1963) proposed a new subgenus, *Tairearea,* to accommodate the New Zealand plants.

Moore and Edgar (1970) discussed the uncertain affinities of *Y. australis*. These authors suggested that the inflorescence was unlike that of other *Yoania*. Instead, they noted features of floral morphology that suggested affinities with the tribes Cranichideae (i.e., *Goodyera* R. Br., *Spiranthes* Rich.) and Neottieae (e.g., *Epipactis* Zinn, *Neottia* Guett.). Yet it was not until 25 years later that *Danhatchia* was established to accommodate plants assigned to *Y. australis* within the Cranichideae (Garay & Christensen, 1995). More recently, specimens morphologically similar to *D. australis* have been discovered in New South Wales, Australia (Garnock-Jones, 2014). First found near Port Macquarie in 2009, additional specimens were discovered at Bundanoon in 2012 (Banks, 2012; Steenbeeke, 2012). Although initially assigned to *D. australis*, Jones and Clements (2018) have suggested these Australian specimens represent a new species, *Danhatchia novaehollandieae* D.L.Jones & M.A.Clem. *tax. nov.*

In New Zealand, *D. australis* occurs on both the North and South Islands. However, it has been reported from just one site on the South Island (i.e., Kaihoka Lakes, Nelson) and in the North Island, it is most commonly found north of Waiuku (de Lange, 2019). Currently, this species is considered "At Risk – Naturally Uncommon" under the New Zealand Threat Classification System (de Lange et al.,

2017). It is not unusual for mycoheterotrophic species to be rare or uncommon in their native habitats, perhaps in part because these plants are easily overlooked (Merckx, 2012; Merckx et al., 2009). At this stage, it remains uncertain whether *Danhatchia* is uncommon or frequently goes unnoticed.

In New Zealand, *D. australis* is sometimes mistaken for a member of *Gastrodia* R. Br., a distantly related genus of terrestrial, mycoheterotrophic orchids with a similar growth habit. However, *Danhatchia* is distinguished from *Gastrodia* by shorter, often glandular-pubescent stems, more numerous scale leaves, and smaller flowers (de Lange, 2019). *Danhatchia* is currently placed in the tribe Cranichideae and subtribe Goodyerinae, on the basis of flower structure, which resembles those of the *Goodyera* alliance (Garnock-Jones, 2014). In particular, floral similarities suggest a relationship with the genus *Gonatostylis* Schltr., a New Caledonian endemic (Pridgeon et al., 1999). However, *Danhatchia* differs from other Goodyerinae in the notable absence of leaves and in some aspects of the floral structure (e.g., a naviculate, or boat-like, labellum) (de Lange, 2019). The placement of *Danhatchia* in the Cranichideae, Goodyerinae also poses an interesting biogeographical question because this subtribe is uncommon in New Zealand and New South Wales. In fact, no other members of the Goodyerinae subtribe have been reported from New Zealand, where *Danhatchia* appears to be most common (Moore & Edgar, 1970). Outside of New Zeland and Australia the Goodyerinae occupy cool, humid habits in Asia, Africa and the Americas (Juswara, 2010). Ultimately, the phylogenetic affinities of *Danhatchia* remain uncertain.

Danhatchia individuals are, like many other holomycoheterotrophs, small, perennial herbs with predominantly subterranean vegetative bodies. In *Danhatchia*, the subterranean portion is a rhizome that is 3-5 mm in diameter and up to 60 cm long, hairy, and often many-branched. Typically, the rhizome is deeply buried beneath decomposing leaf material (de Lange, 2019; Pridgeon et al., 1999). Only the inflorescence is commonly observed above ground, emerging in December to February and standing up to 20 cm tall. Plants lack obvious leaves, although there are vestigial scale leaves along the rhizome. The inflorescence is unbranched and typically bears one to five flowers (de Lange, 2019).

Danhatchia is non-green and individuals are instead pale brown, rose pink, or grey in colour (de Lange, 2019). Although not yet examined empirically, the lack of green pigmentation implies *Danhatchia* is achlorophyllous. As in some other

holomycoheterotrophs, *Danhatchia* may retain low levels of chlorophyll, and hence, some ability to photosynthesise (Merckx, 2012; Suetsugu et al., 2018). However, other characteristics of the genus, such as a lack of obvious leaves and the limited duration of aboveground growth, reinforce the idea that *Danhatchia* is non-photosynthetic. It is certainly hard to imagine how these plants could photosynthesise enough to fully support themselves.

The fungal associate of *D. australis* is thought to be *Lycoperdon perlatum* Pers., the common puffball. Hyphae of this fungus have been observed penetrating the rhizome of *D. australis* (Campbell, 1972). Consistent with this, *D. australis* typically occurs in forests dominated by nikau (*Rhopalostylis sapida*, Arecaceae) and taraire (*Beilschmiedia tarairi,* Lauraceae). Both of these native trees are thought to have a mycorrhizal association with *L. perlatum* (de Lange, 2019).

Currently available mycoheterotrophic plastome sequences represent speciose genera comprised entirely of mycoheterotrophs (e.g., *Thismia* Griff., Burmanniaceae) or dominated by photosynthetic species but with one or more mycoheterotrophic members (e.g., *Corallorhiza* Gagnebin*,* Orchidaceae; *Exochaenium* Griseb., Gentianaceae; *Neottia*). However, many mycoheterotrophic plant genera are small, containing at most a few species, and taxonomically isolated (Merckx, 2012). Although plastomes are yet to be sequenced for most of these groups, they may provide important insights. For example, the extreme morphological reduction of *D. australis*, combined with its taxonomic and geographical isolation, suggests that this species has a long history as a mycoheterotroph. It is therefore reasonable to expect the plastome of *D. australis* to be extensively degraded. Additionally, it is expected that degradation of the *D. australis* plastome will follow the functionally based stepwise pattern of gene loss suggested by the Barrett and Davis (2012) model.

The present study investigates plastome evolution in *D. australis*. The complete plastome sequence of a single New Zealand accession is reported, along with phylogenetic and comparative analyses. The phylogenetic analyses use a broad, but limited sample of Cranichideae and Epidendroideae. These were used to place *D. australis* within a broader orchid phylogeny and therefore identify appropriate species for comparative analyses. Comparative analyses of chloroplast genomes from Goodyerinae were then used to characterise plastome evolution in *D. australis*.

3.2 Materials and Methods

3.2.1 Taxon Sampling

For this study an accession of *D. australis* was collected from Atuanui-Mount Auckland, Auckland (WELT SP107837).

3.2.2 Extraction of total cellular DNA

DNA extractions were conducted using NucleoSpin Plant II kits (Macherey-Nagel) following the manufacturer's instructions. The concentration of extracted DNA was quantified using a Qubit 2.0 Fluorimeter (LifeTechnologies) following the manufacturer's instructions.

3.2.3 Illumina short-read sequencing

The Illumina TruSEQ DNA Nano library preparation methods were used to prepare *D. australis* libraries for sequencing. Briefly, TruSEQ DNA Nano library preparation involves shearing the DNA by ultrasonication into random fragments, with barcoded adapters added to each end during PCR enrichment. The Massey Genome Service (MGS) performed the library preparation and paired-end DNA sequencing on Illumina MiSeq instruments. Following sequencing, the sequence reads were quality assessed using a standard SolexaQA-based workflow (Cox et al., 2010).

3.2.4 Chloroplast genome assembly

Whole chloroplast genome sequences were assembled from quality assessed Illumina 250 bp sequence reads using a genome skimming approach. Next-generation sequencing of total cellular DNA produces reads from all three plant genomes – nuclear, chloroplast and mitochondrial. Genome skimming allows reads from the chloroplast genome to be isolated from the mixture of reads. The chloroplast genome can therefore be assembled without first isolating the chloroplast DNA (Straub et al., 2012; Twyford & Ness, 2017). The custom bioinformatics pipeline used here combined idba_ud v. 1 (Peng et al., 2012), BLAST + 2.3 (Altschul et al., 1990), Geneious R9 (Kearse et al., 2012) and BWA v. 0.7 (Li & Durbin, 2009), as described below.

First, quality assessed reads were assembled into contigs using idba_ud, an iterative de Bruijn graph assembler for paired-end reads that is particularly useful

when coverage of the genome is likely to be uneven (Peng et al., 2012). The resulting contigs were then filtered to remove those that had low similarity to a reference collection of 70 publicly available, complete Orchidaceae plastome sequences that included both photosynthetic and non-photosynthetic species (Appendix I). Custom scripts were used to first remove contigs shorter than 150 nt from the idba_ud results and then to automate BLAST searches against the reference collection. Filtered contigs were then assembled into larger scaffolds using the reference-based and *de novo* assembly tools implemented in Geneious. Reference-based assemblies used plastomes from the photosynthetic *Goodyera procera* (Ker Gawl.) Hook. (Yu, Liang, & Wu, 2015) and *Sobralia callosa* L. O. Williams (Kim et al., 2015).

The IR boundaries were then located by first identifying contigs and scaffolds that contained genes commonly found at these boundaries (e.g., photosystem II protein D1 (*psbA*) and hypothetical chloroplast RF1 (*ycf1*)). Typically, contigs and scaffolds in this set were composed of two distinct sequence portions; one with high identity to all members of this set, the other with high identity to only some members of the set. The IR boundaries were assumed to occur at the point at which identity shifted between these two situations. To complete the circular plastome a second copy of the IR was then manually copied into the sequence.

As a final check, the original Illumina short sequence reads were mapped directly to the draft genome BWA and visualised in Tablet (Milne et al., 2010). Inconsistencies in read coverage were assumed to indicate assembly errors and the draft modified as necessary to correct these.

3.2.5 Chloroplast genome annotation

The draft genome was annotated using tools implemented in Geneious R9 with the *S. callosa* and *G. procera* plastomes as references. To capture the largest number of putative gene regions, initial searches used a low similarity threshold (i.e., 35%).

For putative protein coding genes, the coding sequence was translated and the expected product inspected by eye for appropriate start and stop codons as well as overall similarity, in terms of both length and amino acid sequence, to existing sequences. Each of the putative protein coding regions was then classified as either a functional gene, a pseudogene, or a gene remnant based upon similarity to existing, putatively functional gene sequences.

A putative protein coding region was considered functional if, compared to presumably functional copies of the corresponding gene from other species, the *D. australis* sequence differed by <10% in length and the amino acid positive identity, a description of the extent to which amino acid substitutions disrupt the physiochemical properties (e.g., charge or hydrophobicity) of a protein sequence, was at least 75% based on the BLOSUM62 matrix. Additionally, where functional domains or residues had been identified, these also needed to be retained for the gene to be considered a functional copy. A putative coding region was classed as a pseudogene if it remained recognisable as the corresponding gene, but the amino acid translation suggested it was unlikely to be functional (e.g., sequence was truncated or missing functional domains or residues). Specifically, a putative coding sequence from *D. australis* was assumed to be a pseudogene when the sequence was considerably shorter than presumably functional copies of the corresponding gene (i.e., 21-40% shorter), but retained high amino acid identity (i.e., >90%) or when the sequence was more complete in terms of length (i.e., 11-20% shorter) but amino acid identity was lower (i.e., 60-80%). Finally, a putative coding sequence was considered to be a gene remnant when the sequence was short (i.e., 41-61% shorter) but had high amino acid identity (i.e., >80%) with presumably functional copies of a gene, or when the sequence was longer (i.e., 0-40% shorter) but had lower identity (i.e., 60-89%).

The low similarity threshold used for the initial search resulted in annotations for many putative transfer RNA (tRNA) and ribosomal RNA (rRNA) genes. Annotations were checked manually and where similarity scores were <80% the putative tRNA and rRNA were removed.

3.2.6 Phylogenetic analyses

To evaluate whether members of the Goodyerinae were an appropriate comparison for genome level analyses, separate and combined phylogenetic analyses of three genetic loci were conducted. The loci were selected because they had been useful in previous phylogenetic analyses above species level and were available for the groups of interest; specifically, the Cranichideae (Cranichidinae, Goodyerinae, and Spiranthinae) and Epidendroideae (Caylpsoineae and Neottieae). From the chloroplast genome, portions of the genes encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) and maturase K (*matK*) were included,

as were the nuclear ribosomal internal transcribed spacer region (nrITS, including the ITS1 and ITS2 spacers and 5.8S ribosomal gene). Sequences for the two chloroplast regions were extracted from the complete *D. australis* chloroplast genome and were included in the data set along with published sequences from representatives of the groups of interest (Table 3.1). For the nrITS, a sequence for *D. australis* was assembled by mapping the quality assessed Illumina short sequence reads to the corresponding sequence for *G. procera*; again, this sequence was included in the data set along with published sequences from representatives of the groups of interest (Table 3.1).

Multiple sequence alignments were generated for each of the loci using ClustalO (Larkin et al., 2007) and edited in Geneious R9 to remove portions of the alignment that were ambiguous or where less than 50% of the accessions were represented. For each locus, the best-fit substitution model was identified using the Bayesian Information Criterion (BIC; Schwarz, 1978) as implemented in JModelTest 2.2 (Guindon & Gascuel, 2003; Posada, 2008). Using PhyML (Guindon et al., 2010) and the best-fit model, maximum likelihood (ML) trees as well as ML bootstrap support (500 replicates) was estimated for each locus. A combined data set was compiled by concatenating the four individual matrices and analysed in the same way.

Table 3.1 Details of publicly available DNA sequences used for phylogenetic analyses

Species and authority	Geographical distribution	Genbank accession number		
		matK	*rbcL*	*nrITS*
Corallorhiza mertensiana Bong. [a]	North America	NC_025661	NC_025661	KX755383
Cranichis fertilis (F. Lehm. & Kraenzl.) Schltr	South America	AJ310013	AF074137	AJ000137
Cremastra appendiculata (D.Don) Makino	Asia	NC_037439	NC_037439	KX755389
Dactylostalix ringens Rchb. f.	Asia	KM526761	KM526772	KM526765
Epipactis mairei Schltr.	Asia	MG925367	MG925367	KF419101
Gonatostylis vieillardii Rchb. f.	New Caledonia	AJ310034	AY381122	FJ473325
Goodyera procera (Ker Gawl.) Hook.	Asia	NC_029363	NC_029363	JN114516
Goodyera schlechtendaliana Rchb. f.	Asia	NC_029364	NC_029364	KR815835
Kuhlhasseltia nakaiana (F. Maek.) Ormerod	Asia	KY354041	KY354041	KT338723
Ludisia discolor (Ker Gawl.) A. Rich.	Asia	NC_030540	NC_030540	KT344102
Neottia ovata (L.) Bluff & Fingerh.	Europe, Asia	NC_030712	NC_030712	FJ694841
Neottia pinetorum (Lindley) Szlach.	Asia	NC_030710	NC_030710	KT338759
Spiranthes cernua (L.) Rich.	North America	AF263682	AF074229	MF170212
Spiranthes sinensis (Pers.) Ames	Asia, Pacific	JF972946	JF972913	MF286510
Yoania japonica Maxim. [a]	Asia	–	–	LC176609

a – species which are non-photosynthetic

3.2.7 Comparative analyses

Complete plastome sequences for eight photosynthetic Goodyerinae were recovered from GenBank (Table 3.2).

Gene and protein sequence similarity — Using both nucleotide and inferred protein sequences, multiple sequence alignments were constructed for all putatively functional protein-coding genes from the Goodyerinae plastomes using ClustalO. In cases where the published gene annotations differed in terms of start and stop codons, alignments were trimmed to the most common version where this preserved the reading frame. For amino acid sequences that included a premature stop codon, the sequence beyond this point was considered for similarity if the reading frame was otherwise intact and amino acid identity remained high (i.e., >70%).

In a few cases, genes not included in the published annotation were identified upon closer inspection. Typically, these were genes that consist of multiple exons or have non-standard start codons (i.e., may undergo RNA editing). These situations may have confounded automated tools for annotation transfer or open reading frame (ORF) detection. For example, there may be too few residues in very short exons to meet the similarity threshold even if identity with the reference is 100%. As a result, sequences for the cytochrome b6 (*petB*) gene from *Goodyera schlectendaliana* Rchb. f. and *G. procera*, the ribosomal protein L2 (*rpl2*) gene from *G. schlechtendaliana* and the photosystem II CP47 (*psbB*) gene from *Ludisia discolor* (Ker Gawl.) A.

Table 3.2 Details of publicly available complete Goodyerinae plastome sequences

Species and authority	Geographic distribution	Genbank accession number	Reference
Anoectochilus emeiensis K.Y Lang	Asia	LC057212	Zhu et al. (2016)
Anoectochilus roxburghii (Wall.) Lindl.	Asia	KR779936	Zeng & Guo (direct submission)
Goodyera fumata Thwaites	Asia	KJ501999	Lin et al. (2015)
Goodyera procera (Ker Gawl.) Hook.	Asia	NC_029363	Yu, Lian & Wu (direct submission)
Goodyera schlechtendaliana Rchb. f.	Asia	NC_029364	Yu, Lian, & Wu (direct submission)
Goodyera velutina Maxim. ex Regel	Asia	KT886432	Yu., Lian, & Wu (direct submission)
Kuhlhasseltia nakaiana (F. Maek.) Ormerod	Asia; New Guinea	KY354041	Kim & Kim (direct submission)
Ludisia discolor (Ker Gawl.) A. Rich.	Asia	NC_030540	Yu & Wu (direct submission)

Rich. were included in our analyses, despite not being present in published annotations.

Summary statistics (e.g., length, nucleotide and amino acid identity) were calculated for each gene alignment. Two-sample t-tests (Snedecor & Cochrane, 1989) were used to compare variation in gene length and identity between genome regions (i.e., LSC, SSC, and IRs) and gene classes based on those identified in the Barett and Davis (2012) model.

Repeats — The Geneious repeat finder tool was used to locate repeats in the LSC and SSC regions of each plastome; repeats unique to the IR were excluded from the analysis. Two searches were carried out on each plastome. The first identifed repeats of >30 nt in length; the second those 20-29 nt in length. Preliminary searches for smaller repeats (10-19 nt in length) consistently resulted in large numbers per genome (e.g., >10,000) and these were not analysed further. Only exact repeats were considered and repeats that were wholly enclosed within another larger repeat were also excluded.

3.3 Results

3.3.1 DNA Sequencing and genome assembly

The plastome of a single New Zealand *D. australis* specimen was assembled and annotated (Figure 3.1). This sequence has been deposited in Genbank with the accession number MN251184. Summary statistics for the DNA sequencing and plastome assembly are presented in Table 3.3.

3.3.2 Phylogenetic analysis

Maximum likelihood analyses of the nrITS, chloroplast, and combined matrices all resulted in highly similar phylogenetic trees (Figure 3.2). In each case, four main clades were recovered; these clades are consistent with the current taxonomy of the included species. Specifically, there are clades corresponding to the Cranichideae, subtribe Goodyerinae (clade I), the Cranichideae, subtribes Spiranthinae and Cranichidinae (clade II), the Epidendroideae, subtribe Neottiae (clade III), and the Epidendroideae, subtribe Calypsoineae (clade IV). These four clades were strongly

supported in all three analyses (bootstrap support [bs] = 80-100), as was the split between the Cranichideae (clades I and II) and the Epidendroideae (clades III and IV) (bs = 100). *Danhatchia australis* falls within clade I in analyses of all three data sets (bs >100). Within this clade, resolution and support for relationships is good, with *D. australis* always strongly supported as sister to the pairing of *Kuhlhasseltia nakaiana* (F. Maek.) Ormerod and *L. discolor* (bs >90%). Statistics for the matrices and phylogenetic analyses are presented in Table 3.4.

3.3.3 Overall genome size and structure

The plastome of *D. australis* has the four-part structure typical of most photosynthetic angiosperms and is only slightly smaller than those available for photosynthetic Goodyerinae (Table 3.5). For the photosynthetic Goodyerinae, the plastome ranges from 147,614-156,252 nt long, whereas that of *D. australis* is 147,326 nt in length.

The IR, SSC and LSC regions of the *D. australis* plastome are also broadly consistent with those of photosynthetic Goodyerinae in terms of size (Table 3.5). Among the photosynthetic species, the IRs range in size from 26,162 to 26,612 nt, the LSC regions from 81,617 to 84,077 nt, and the SSC regions from 13,673-18,342 nt. For *D. australis* the IRs and SSC region are 506 and 379 nt shorter, respectively, than those of *K. nakaiana*, which is otherwise the smallest plastome yet reported for a representative of Goodyerinae. In contrast, the *D. australis* LSC region is 103 nt larger than that of *K. nakaiana*.

Table 3.3 Summary statistics for *Danhatchia australis* (Hatch) Garay & Christenson chloroplast genome assembly

Number of Illumina short sequence reads	6,895,744
Number of Illumina short sequence reads mapped to plastome	68,037
% of reads mapped to plastome	1%
Mean and, in parentheses, range of read coverage	
Overall	114.3 (11-296)
Inverted Repeat (IR)	161.9 (34-296)
Small single copy (SSC) region	57.7 (11-159)
Large single copy (LSC) region	93.2 (11-240)

Figure 3.1 The plastome of *Danhatchia australis*. The plastome size, structure, and overall gene content are all similar to available plastomes from Goodyerinae, with the exception of the loss of *ndh* genes from the SSC region. Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).

Figure 3.2 Maximum likelihood phylogenies for the nrITS, chloroplast, and combined DNA sequence matrices. A, nrITS; B, chloroplast (combined maturase K (*matK*) and ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) sequences); and C, combined. In all three phylogenetic trees, clade I corresponds to the Cranichideae, subtribe Goodyerinae; clade II, Cranichideae, subtribes Spiranthinae and Cranichidinae; clade III, Epidendroideae, subtribe Neottieae; and clade IV, Epidendroideae, subtribe Calypsoineae. Numbers associated with branches are bootstrap values. Full taxon names are listed in Table 3.1.

Although the plastome of *D. australis* is broadly similar in size to those available for the Goodyerinae, the plastome to which it is closest, that of *K. nakiana*, is markedly smaller than the other Goodyerinae plastomes. Specifically, the IR, LSC and SSC regions of *K. nakiana* are 157, 415, and 3560 nt smaller, respectively, than the photosynthetic Goodyerinae plastome next closest in size for the corresponding region (Table 3.5).

Considered in relation to overall plastome size, the sizes of the IR, LSC and SSC regions of *D. australis* and *K. nakiana* differ consistently from those of the remaining Goodyerinae. In these two genomes the IR (i.e., >35.45% c.f. <34.75%) and LSC (i.e., >55.29% c.f. <54.80%) regions are larger but the SSC region is smaller (i.e., <9.26% c.f. >11.25%). This implies that smaller overall plastome size in *D. australis* and *K. nakaiana* is associated with a reduced SSC region. The boundaries between the IR and LSC regions are broadly consistent across all nine Goodyerinae plastomes (Figure 3.3). The boundary with IR$_B$ is always located within the ribosomal protein L22 gene (*rpl22*), although the boundary position within the gene differs between species; specifically, from nucleotide position 44 in *Anoectochilus roxburghii* (Wall.) Lindl. to nucleotide position 181 in *D. australis*. The boundary between the LSC region and IR$_A$ is always within the intergenic spacer between the ribosomal protein S19 (*rps19*) and *psbA* genes. Again, the boundary position differs between species, being 232-384 nt from the terminus of *rps19* and 75-112 nt from the start of *psbA*.

The boundaries between the IRs and SSC are also broadly consistent. The boundary with IR$_B$ is located either within *ycf1* or close to the terminus of this gene. The boundary with IR$_A$ has the same position relative to *ycf1*, but in this case also often overlaps with the NADH dehydrogenase subunit 5 (*ndhF*) gene that is

Table 3.4 Statistics for matrices and phylogenetic analyses

	matK	*rbcL*	chloroplast	nrITS	Combined
Matrices					
No. of taxa	15	15	15	16	16
Matrix length (nt)	1564	1454	3028	692	3720
No. of varied sites	402	160	562	440	1002
Maximum Likelihood analysis					
Log likelihood	–	–	-8982.201	-5187.125	-14630.579
Substitution model	–	–	GTR + I + G	SYM + I + G	GTR + I + G

otherwise located within the SSC region. Expansion of the SSC region in
K. nakaiana and *D. australis* is due to this boundary being shifted relative to the
remaining species. In these two species, the IR boundary is closer to the start of *ycf1*
such that more of this gene falls within the SSC region.

3.3.4 GC content

The GC content was similar for all available Goodyerinae plastomes, including that
of *D. australis* (Table 3.5). The overall GC content falls between 36.90 and 37.70%
for these genomes; GC content does not appear to be phylogenetically structured as
Anoectochilus Blume species occur at both ends of this range and GC content varies
between *Goodyera* species. The overall GC content for *D. australis*, 37.20%, is also
the mean GC content for the photosynthetic plastomes.

  The IR, LSC, and SSC regions of the photosynthetic plastomes differed in
terms of GC content. Specifically, GC content was higher for the IR (43.00-43.50%)
than either of the single copy regions; moreover, GC content was higher for the LSC
region (34.30-35.50%) than the SSC region (28.50-31.20%). As for GC content
overall, that of the IR, LSC, and SSC regions in *D. australis* were consistent with
those for the photosynthetic plastomes.

3.3.5 Coding and non-coding sequences

The available photosynthetic Goodyerinae plastomes, as well as that of *D. australis*,
all had similar proportions of coding and non-coding DNA. Specifically, putatively
functional genes occupied 47-53% of the sequence length with a further 0-10%
occupied by pseudogenes or gene remnants. Non-coding DNA, including intergenic
spacer regions and introns, makes up 47-48% of the genome (Table 3.6).
The proportion of the plastome occupied by any one of these three sequence
categories appears not to be strongly linked to plastome size. That said, the two
smallest genomes, those of *D. australis* and *K. nakaiana*, have the lowest proportion
of putatively functional genes and their proportions of disrupted genes and non-
coding DNA tends to be towards the upper limit of the typical range for Goodyerinae
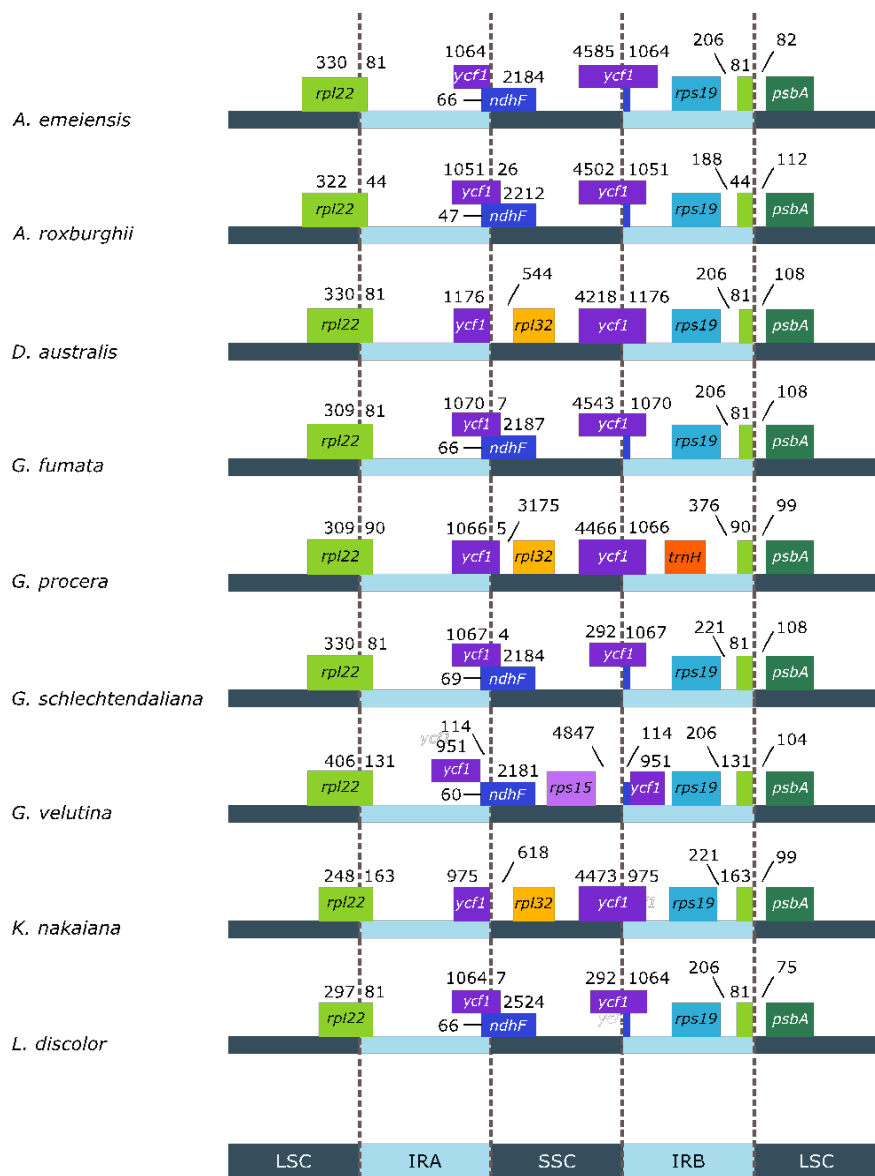(Table 3.6).

Figure 3.3 Positions of the boundaries between the inverted repeats (IR) and single copy (SC) regions in available Goodyerinae plastomes. The IR boundaries are well-conserved in this subtribe. The LSC-IR$_A$ boundary is always within the gene for ribosomal protein L22 (*rpl22*), whereas the LSC-IR$_B$ boundary is always near the start of the gene for photosystem II protein D1 (*psbA*). The SSC-IR$_A$ boundary is within the gene for NADH dehydrogenase subunit 5 (*ndhF*), if a functional copy of this gene is present, or is within the ribosomal L32 (*rpl32*)-*trnN* intergenic spacer. The SSC-IR$_B$ boundary is also within or near the terminus of the gene hypothetical chloroplast RF1 (*ycf1*).

Table 3.5 Summary statistics for the Goodyerinae plastome sequences

Species	Plastome		IR			SSC			LSC		
	Size (nt)	GC content (%)	Size (nt)	% plastome	GC content (%)	Size (nt)	% plastome	GC content (%)	Size (nt)	% plastome	GC content (%)
Anoectochilus emeiensis	152,650	36.90	26,319	34.48	43.20	17,342	11.36	29.30	82,670	54.16	34.50
Anoectochilus roxburghii	156,252	37.70	26,591	34.04	43.20	18,405	11.78	31.2	84,665	54.18	35.70
Danhatchia australis	147,326	37.20	25,656	35.07	43.30	13,294	9.09	29.40	81,720	55.84	34.7
Goodyera fumata	155,643	37.30	26,612	34.17	43.30	18,342	11.78	29.90	84,077	54.02	35.1
Goodyera procera	153,240	37.60	26,401[a]	34.46[a]	43.50[a]	18,406	12.09	30.00	82,032	53.53	35.5
Goodyera schlechtendaliana	154,348	37.20	26,541	34.39	43.30	18,051	11.70	29.70	83,215	53.19	34.9
Goodyera velutina	152,692	36.90	26,495[b]	34.70[b]	43.20[b]	17,259	11.30	29.30	82,443	53.99	34.5
Kuhlhasseltia nakaiana	147,614	36.90	26,162	35.45	43.00	13,673	9.26	28.50	81,617	55.29	34.3
Ludisia discolor	153,054	37.00	26,573[c]	34.72[c]	43.10[c]	17,233	11.26	29.70	83,861	54.79	34.6

[a] The regions designated as the IR in the publically available *G. procera* chloroplast genome sequence are not identical; two nucleotide substitutions are inferred between the IR copies, these may represent errors (i.e., in sequencing or assembly) or variation within or between chloroplasts. This corresponds to < 0.1% difference and these regions are therefore considered to be the IRs for the purposes of this comparison.

[b] The regions designated as the IR in the publically available *G. velutina* chloroplast genome sequence are not identical; 10 nucleotide substitutions and two indels (i.e., 1 nt differences in the length for thymine homopolymer repeats) are inferred between the IR copies, these may represent errors (i.e., in sequencing or assembly) or variation within or between chloroplasts. This corresponds to < 0.1% difference and these regions are therefore considered to be the IRs for the purposes of this comparison.

[c] The regions designated as the IR in the publically available *Ludisia discolor* chloroplast genome sequence are not identical; 23 nucleotide substitutions are inferred between the IR copies, these may represent errors (i.e., in sequencing or assembly) or variation within or between chloroplasts. This corresponds to < 0.1% difference and these regions are therefore considered to be the IRs for the purposes of this comparison

3.3.6 Gene content

Gene content is broadly similar across all the Goodyerinae plastomes including that of *D. australis*. The plastome of *Anoectochilus emeiensis* K.Y Lang has the largest number of functional genes (107) with the remainder having subsets of these. All the remaining plastomes retain full complements of the *rpo* subunit, ribosomal RNA and transfer RNA genes. All the housekeeping genes are also retained in most of the remaining plastomes; the exceptions involve *ycf1*, which is truncated in *G. schelechtendaliana*, *G. velutina* and *L. discolor*. Five of the Goodyerinae plastomes have lost one to five genes (Table 3.7). For example, in *A. roxburghii* and *Goodyera fumata* Thwaites the hypothetical chloroplast RF15 (*ycf15*) has been pseudogenised. The plastomes of *D. australis*, *G. procera*, and *K. nakaiana* each have 10 genes that are either disrupted or have been lost (Table 3.7). In both *D. australis* and *K. nakaiana* these disruptions and absences mostly involve genes encoding subunits of the NADH dehydrogenase complex. The exception is the cytochrome c biogenesis (*ccsA*) gene, which in *D. australis* is truncated. In *G. procera*, only one NADH dehydrogenase subunit gene is disrupted, with most disruptions involving photosystem II or ribosomal protein genes.

Although functionally lost, the coding regions of most of the disrupted genes remained identifiable due to high nucleotide and amino acid identity with putatively functional genes. Disruption of the genes is commonly caused by short indels, 1-10 nt in length, that interrupt the reading frame and introduce premature stop codons. Alternatively, gene disruption was the result of a premature stop codon introduced as a result of a nucleotide substitution.

3.3.7 Gene variability

Putatively functional genes vary little in overall length in the Goodyerinae plastomes, including that of *D. australis*. There is no apparent difference in patterns of length variation between gene classes, location within the genome (i.e., SSC, LSC or IR) or between those genes that are retained by all species and those that have been functionally lost from at least one genome (Table 3.7).

For most genes, the inferred amino acid sequences vary by < 10 amino acids (Table 3.8). There are exceptions with six genes varying by 20-49 amino acids and a further six by > 50 amino acids. Within this latter class, three genes vary by > 100

Table 3.6 Summary statistics for gene content of Goodyerinae plastomes

Species	Plastome size (nt)	Functional genes			No. of pseudogenes	Disrupted genes			Non-coding		No. of missing genes[a]
		No. of genes	Size (nt)	% plastome		No. of gene remnants	Size (nt)	% plastome	Size (nt)	% plastome	
A emeiensis	152,650	107	80,839	0.53	0	0	0	0.00	71,811	0.47	–
A. roxburghii	156,252	106	80,575	0.52	0	1	227	0.00	75,450	0.48	0
D. australis	146,326	97	70,785	0.48	2	3	6,468	0.09	70,073	0.48	5
G. fumata	155,643	106	80,736	0.52	1	0	239	0.00	74,668	0.48	0
G. procera	153,240	97	74,296	0.48	8	1	6,626	0.09	72,318	0.47	1
G. schlechtendaliana	154,348	104	73,225	0.47	2	0	7,348	0.10	73,775	0.48	1
G. velutina	152,692	102	72,307	0.47	2	2	7,997	0.11	72,388	0.47	1
K. nakaiana	147,614	97	70,375	0.48	3	3	7,389	0.10	69,850	0.47	4
L. discolor	153,054	105	74,884	0.49	1	0	5,931	0.08	72,239	0.47	1

[a] The gene content of *A. emeiensis* was used as the basis for calculating the number of missing genes

Table 3.7 Plastome gene content of the Goodyerinae

					Species					
		A. emerensis	*A. roxburghii*	*D. australis*	*G. fumata*	*G. procera*	*G. schlechtendaliana*	*G. velutina*	*K. nakaiana*	*L. discolor*
NADH dehydrogenase genes										
NADH dehydrogenase subunits		11/11 present	11/11 present	2/11 present; *ndhC* and *ndhI* pseudogenes; *ndhD* and *ndhK* remnants; *ndhA*, *ndhF*-*ndhH* and *ndhJ* absent	11/11 present	10/11 present; *ndhF* a remnant	11/11 present	10/11 present; *ndhD* a remnant	1/11 present; *ndhC*, *ndhG* and *ndhH* pseudogenes; *ndhA*, *ndhD* and *ndhI* remnants; *ndhB*, *ndhF*, *ndhJ* and *ndhK* absent	11/11 present
Photosynthesis genes										
Cytochrome		6/6 present	6/6 present	6/6 present	6/6 present	6/6 present	6/6 present	6/6 present	6/6 present	6/6 present
photosystem I		4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	3/4 present, *psaI a* pseudogene	4/4 present	4/4 present
photosystem II		15/15 present	15/15 present	15/15 present	15/15 present	11/15 present; *psbB*, *psbM* and *psbZ* pseudogenes, *psbH* absent	13/15 present; *psbB* and *psbM* pseudogenes	15/15 present	15/15 present	15/15 present

other photosynthesis genes	6/6 present	5/6 present; *ycf15* a remnant	5/6 present; *ccsA* a remnant	5/6 present; *ycf15* a pseudogene	6/6 present	6/6 present	4/6 present; *ycf4* a pseudogene, *cemA* a remnant	6/6 present	6/6 present
RNA polymerase genes									
RNA polymerase subunits	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present
ATP synthase genes									
ATP synthase subunits	6/6 present	6/6 present	6/6 present	6/6 present	5/6 present; *atpF* a pseudogene	6/6 present	6/6 present	6/6 present	6/6 present
ribosomal protein S	12/12 present	12/12 present	12/12 present	12/12 present	9/12 present; *rps12*, *rps16* and *rps18* pseudogenes	12/12 present	12/12 present	12/12 present	11/12 present; *rps12* a pseudogene
ribosomal protein L	9/9 present	9/9 present	9/9 present	9/9 present	8/9 present; *rpl2* a pseudogene	9/9 present	9/9 present	9/9 present	9/9 present
ribosomal RNAs	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present	4/4 present
transfer RNAs	21/21 present	21/21 present	21/21 present	21/21 present	21/21 present	21/21 present	21/21 present	21/21 present	21/21 present
Housekeeping genes	6/6 present	6/6 present	6/6 present	6/6 present	6/6 present	5/6 present; *ycf1* absent	5/6 present; *ycf1* absent	6/6 present	5/6 present; *ycf1* absent

Table 3.8 Statistics for gene variation in Goodyerinae species

	Mean and range (in parentheses) of gene length (nt)	Mean and range (in parentheses) of gene length variation (nt)	Mean and range (in parentheses) of mean nucleotide identity (%)	Mean and range (in parentheses) of mean amino acid identity (%)	Mean and range (in parentheses) of pairwise positives (%)
Genes not disrupted in any of the sampled species	839 (90-6678)	31.8 (0-498)	97.6 (91.0-99.7)	96.8 (86.5-100.0)	98.0 (89.3-100)
Genes disrupted in at least one of the sampled species	908 (105-5559)	29.3 (0-201)	97.2 (89.9-99.5)	95.6 (81.6-99.6)	97.1 (87.1-100)
Genes located in the IR	1991 (219-6549)	73.9 (0-375)	96.9 (89.1-99.7)	96.3 (81.40-99.70)	97.4 (87.30-99.80)
Genes located in the SSC	822 (173-2252)	22.4 (0-159)	96.6 (94.1-97.6)	95.0 (91.0-100)	96.8 (93.20-100)
Genes located in the LSC	724 (90-4100)	27.4 (0-498)	97.5 (91.2-99.6)0	96.5 (86.50-100)	98.0 (89.40-100)
Genes in the ndh gene class	959 (306-2252)	21.8 (0-177)	97.2 (96.0-99.6)	95.1 (86.5-99.6)	97.0 (91.0-99.6)
Genes in the photosynthesis-gene class	581 (90-2253)	9.10 (0-159)	98.0 (94.1-99.6)	97.6 (90.3-100)	98.6 (93.2-100)
Genes in the non-photosynthesis gene class	1060 (114-6768)	51.6 (0-498)	96.9 (89.1-99.8)	95.7 (81.4-100)	97.2 (87.3-100)

amino acids; specifically, the RNA polymerase subunit beta" (rpoC2) gene and the hypothetical protein genes ycf1 and ycf2. However, these three genes are among the longest encoded by the plastome and this variation represents a relatively small proportion of overall gene length. Only four genes retained by D. australis differ in length from photosynthetic Goodyerinae. The NADH dehydrogenase subunit 2, the photosystem I P700 chlorophyll a apoprotein A1 and ycf1 genes are 66-165 nt shorter than the mean for photosynthetic Goodyerinae. The photosystem II reaction center N protein (psbN) gene is 3 nt longer.

When only putatively functional gene sequences from the photosynthetic Goodyerinae are considered, residue identity tends to be higher for genes that are functionally retained across the available Goodyerinae plastomes than for genes that are disrupted in at least one species. However, these differences were not statistically significant (two-sample t-tests, t = -1.40 to -0.33, p-value = 0.171-0.741, DF = 36-39). Likewise, the extent of length variation for any given gene is not correlated with whether the gene is disrupted in one or more of the available Goodyerinae plastomes (two-sample t-test, t = -0.15, p-value = 0.881, DF = 67). For example, the *ccsA* gene is functionally lost from the *D. australis* plastome and varies by 5.8% at the nucleotide level across the remaining eight plastomes, whereas the acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (*accD*) has higher diversity (9.0%) but is retained in *D. australis*. Results were similar for amino acid identity and positive amino acid identity. These observations suggest that underlying levels of variability do not predispose genes to disruption in Goodyerinae.

Results for gene classes were broadly similar. In this case, three groups were considered; the *ndh* genes, the photosynthesis genes, and a group containing the remaining non-photosynthesis genes. In preliminary analyses, differences between the photosynthesis and non-photosynthesis gene classes were statistically significant for nucleotide identity (two-sample t-test, t = -2.59, p-value = 0.012, DF = 60), amino acid identity (two-sample t-test, t = -2.56, p-value = 0.013, DF = 64), and positive amino acid identity (two-sample t-test, t = -2.47, p-value = 0.016, DF = 64). However, these results appeared to be driven by outliers with unusually low identities in the non-photosynthesis gene class. Specifically, *accD*, *ycf1* and the hypothetical chloroplast protein RF2 (*ycf2*) genes had identities < 92% for both nucleotides and amino acids whereas for other genes in this class identities were typically > 93%. These genes are also longer and exhibit greater length variation than

other members of this gene class. In particular, high levels of variability have been reported for *accD* and *ycf2* from photosynthetic species including members of *Ipomoea* L. (Park et al., 2018) and *Byrsonima* Rich ex Kunth (Menezes et al., 2018). With these three genes excluded, no statistically significant differences consistent with the Barrett and Davis model (2012) were found (two-sample t-tests, t = -2.07-0.76, *p*-values = 0.057-0.787, DF = 12-35). Statistically significant differences were only observed in comparisons of amino acid identity involving the photosynthesis gene class (two-sample t-tests, t = -2.04 to -2.29, *p*-value = 0.038-0.045, DF =14-63).

In the plastomes examined, identity at the level of amino acids also did not differ significantly between genes from the IR, LSC or SSC regions (amino acid identity two sample t-test, t = -0.12-1.82, *p*-value = 0.088-0.911, DF = 7-16; positive amino acid identity two sample t-test, t = -0.33-1.71, *p*-values = 0.108-0.749; DF = 7-15). Mean gene lengths (two sample t-tests, t = -0.44-1.35, *p*-values = 0.218-0.666, DF = 7-15) and gene length variation (two sample t-tests, t = 0.30-1.00 *p*-values = 0.346-0.769, DF = 7-21) also do not vary significantly between regions. In contrast, the mean pairwise nucleotide identity of genes does differ significantly between the LSC and SSC regions (two-sample t-test, t = 2.24, *p*-value = 0.040, DF = 16). Comparisons of nucleotide identity involving the IR region were not statistically significant (two sample t-tests, t = -0.42-0.18, *p*-value = 0.689-0.861, DF = 7). These results imply higher rates of substitution for nucleotide, but not amino acid, sequences for genes in the SSC region. This may be explained by differences in the distribution and variability of *ndh* and photosynthesis-related genes. About two thirds of the genes in the SSC region of Goodyerinae plastomes belong to the more variable *ndh* class, whereas most of the photosynthesis genes are in the LSC region.

3.3.8 Repeats

Short repeat sequences are not common in the available Goodyerinae plastomes, which contain 11-25 repeats 20-29 nt long and, with one exception, up to 4 repeats > 30 nt long (Table 3.9). The plastome of *A. roxburghii* contained 14 repeats > 30 nt in length with several clustered within a single low complexity, AT-rich region, between *trnT* and *trnL* in the LSC region. Most repeats are present in two copies.

In Goodyerinae, most repeats > 30 nt in length are palindromes. For repeats 20-29 nt long, the repeat types are split more evenly between palindromes, inverted, and direct repeats. Most repeats are located within intergenic spacers in the LSC

region and, in general, are less than ~80 nt in length. In *A. roxburghii* they are up to 140 nt. The number of repeat sequences is not linked to plastome size and is not taxonomically structured. The plastome of *D. australis* has a comparable number of repeats, similar in size, type, and location to other Goodyerinae. In a few cases, repeats are shared between species. For example, a 46 nt repeat in the intergenic spacer between the photosystem II reaction centre protein M (*psbM*) gene and cytochrome b6-f complex subunit 8 (*petN*) gene is shared by *A. emeiensis*, *D. australis*, *G. fumata*, *K. nakaiana*, and *L. discolor* with *A. roxburghii* having a shorter, 34 nt version of this repeat. *Danhatchia australis* also shares repeats with other Goodyerinae. For example, a 48 nt repeat within the *clpP-psbB* intergenic spacer is shared with *G. fumata* and a 31 nt direct tandem repeat within the *rps19-psbA* intergenic spacer is shared with *G. schlechtendaliana*.

3.4 Discussion

To understand the evolutionary trajectory of plastomes in mycoheterotrophic plant species it is useful to compare them to the plastomes of photosynthetic relatives. To identify appropriate comparisons, phylogenetic analyses of both chloroplast and nuclear markers were conducted. Results of these analyses are described, followed by comparative analyses with members of Goodyerinae.

3.4.1 Phylogenetic analyses

Analyses of the chloroplast, nuclear, and combined datasets strongly suggest that *D. australis* is most closely related to members of the Cranichideae. This is consistent with the current taxonomic arrangement (Figure 3.2). In these analyses, *D. australis* was consistently placed within a clade containing other representatives of this tribe (bs 100% in all three analyses). The present analyses include representatives of three subtribes of the Cranichideae; the Cranichidinae (i.e., *Cranichis* Sw.), the Goodyerinae (e.g., *Goodyera* and *Gonatostylis*), and the Spiranthinae (i.e., *Spiranthes*). Members of each of these groups form distinct lineages in phylogenetic analyses, with strong support (bs >90%) for *D. australis* grouping with members of the Goodyerinae. Within the Goodyerinae, *D. australis* is sister to a clade containing *L. discolor* and *K. nakaiana*, again with strong support (bs

Table 3.9 Size and location of repeats in the Goodyerinae plastome sequences

| Species | Number of repeats | Repeat size class[a] | Location of repeat | | | | | |
| | | | Functional class | | | Genome region | | |
			Intron	Intergenic spacer	Coding sequence	LSC	SSC	IR
A. emeiensis	15	20-29	3	9	3	12	1	2
	4	>30 (31-46)	0	4	0	4	0	0
A. roxburghii	18	20-29	0	14	4	16	0	2
	14	>30 (40-140)	1	12	1	9	5[b]	1[b]
D. australis	11	20-29	0	7	4	10	1[b]	1[b]
	6	>30 (31-50)	1	3	2	6	0	0
G. fumata	25	20-29	5[e]	18[e]	5	22	0	3
	3	>30 (34-48)	0	3	0	2	1	0
G. procera	18	20-29	1	14	3	17[f]	1[f]	0
	0	–	0	0	0	0	0	0
G. schlechtendaliana	18	20-29	3[e]	14[e]	2	16[f]	0	3[d]
	3	>30 (31-48)	1	2	0	3	0	0
G. velutina	15	20-29	4[e]	9[e]	3	13[f]	2[f]	0
	2	>30 (40-83)	0	1	1	2	0	0
K. nakaiana	11	20-29	2[c]	6	4[c]	1[d]	0	1[d]
	2	>30 (40-46)	0	2	0	1	1	0
L. discolor	15	20-29	3[e]	10[e]	3	14	1[b]	2[b]
	1	>30	0	1	0	1	0	0

[a] For repeats greater than 30 nt in length the observed size range of the repeats is given in parentheses.

[b] One copy of one repeat is located within the SSC region and the other in the IR.

[c] One copy of one repeat is located within an intron and the other in a coding sequence.

[d] One copy of one repeat is located within the LSC region and the other in the IR.

[e] One copy of one or more repeats is located within an intron and the other in an intergenic spacer.

[f] One copy of one repeat is located within the LSC region and the other in the SSC region.

100% in all three analyses). Although sampling remains limited, these results suggest lineages in phylogenetic analyses, with strong support (bs >90%) for *D. australis* grouping with members of the Goodyerinae. Within the Goodyerinae, *D. australis* is sister to a clade containing *L. discolor* and *K. nakaiana*, again with strong support (bs 100% in all three analyses). Although sampling remains limited, these results suggest that *D. australis* is not merely a leafless, achlorophyllous member of a larger and more widespread genus, such as *Goodyera* or *Spiranthes.*

3.4.2 Comparative genome analyses

The plastome of *D. australis* was compared to seven publically available complete plastome sequences representing four genera – *Anoectochilus*, *Ludisia*, *Goodyera* and *Kuhlhasseltia* – of the Goodyerinae. The *D. australis* plastome is broadly similar in size, structure and gene content to these plastomes, all of which are from photosynthetic species.

  The *D. australis* plastome is 1288 nt smaller than that of the next largest Goodyerinae plastome (i.e., *K. nakaiana*) and 9926 nt smaller than the largest of the available plastomes (i.e., *A. roxburghii*). These differences are relatively minor compared to those of other holomycoheterotrophs. For example, in *Corybas* Salisb. (Orchidaceae) the plastome of the sole holomycoheterotrophic representive, *C. cryptanthus* Hatch, is ~65,000 nt smaller than those of its photosynthetic relatives (Chapter 2). The reduced size of the *D. australis* genome reflects smaller IR and SSC regions (Table 3.5). Again these changes are relatively minor compared to some mycoheterotrophic species. For example, there have been dramatic changes in IR length in *Epipogium* J.G.Gmel. ex Borkh. ((Schelkunov et al., 2015)). The IR boundary positions in the *D. australis* plastome are also broadly similar to those in the other available Goodyerinae plastomes (Figure 3.3). Although exact nucleotide position of the boundaries differ, they are located within or between the same genes as in the photosynthetic species.

  In *D. australis* and *K. nakaiana*, smaller plastome sizes appear to be correlated with loss of many of the NADH dehydrogenase subunit genes (Table 3.7). This would be consistent with the smaller SSC region in *D. australis*, as many of the *ndh* genes are located within this portion of the plastome in the photosynthetic Goodyerinae. Aside from the *ndh* genes, the only other gene loss from either of these

two species is *ccsA* in *D. australis*. Loss of *ccsA* is unique to this species and this gene is putatively functional in the other Goodyerinae plastomes.

Goodyerinae plastomes have similar numbers of functional genes. That said, a variety of protein coding genes have been lost from the plastomes of photosynthetic Goodyerinae. In particular, *G. procera* remains green, and presumably photosynthetic, despite losing the same number of genes as *D. australis*. The other genome characteristics (e.g., repeat content, GC% and proportion of non-coding and coding DNA) were also very similar between the plastome of *D. australis* and the photosynthetic Goodyerinae.

3.4.3 Plastome evolution

The phylogenetic analyses place *D. australis* sister to a pairing of *L. discolor* and *K. nakaiana*. Consistent with this, the plastomes of *D. australis* and *K. nakaiana* are more alike than those available for other members of the Goodyerinae. Specifically, the plastomes of *D. australis* and *K. nakaiana* are smaller than those of the others examined. Reduced overall plastome size in these two species is linked to a smaller SSC region due to loss of the *ndh* genes, the majority of which fall within this region in photosynthetic relatives. However, comparative analyses suggest that these similarities are likely to reflect parallel evolution rather than shared phylogenetic history. For example, although most *ndh* genes are disrupted in both plastomes, the number of intact *ndh* genes differs between them, as does the number and identity of *ndh* genes that are pseudogenes, remnants or absent (Table 3.7).

It appears that broadly similar, although not identical, changes have occurred in the plastomes of both *D. australis* and *K. nakaiana*. This implies that these changes are not themselves responsible for the transition to a mycoheterotrophic lifestyle in *D. australis*. Other examples of mycoheterotrophic plant species with plastomes that differ little from those of their photosynthetic relatives include members of *Corallorhiza*, *Cymbidium* Swartz (Orchidaceae), and *Exochaenium*. In *Cymbidium*, the sole non-photosynthetic species, *Cymbidium macrorhizon* Lindl. is sister to the photosynthetic *Cymbidium lancifolium* Hook. (Kim et al., 2018). The plastomes of these two species differ by less than 100 nt (149,859 nt and 149,945 nt, respectively) but are ~300-7500 nt smaller than other photosynthetic *Cymbidium*. With the exception of the *ndh* genes, of which they retain only *ndhC,* and *ycf15*, *C. macrorhizon* and *C. lancifolium* both have a full gene complement (Kim et al.,

2018). Photosynthetic capacity in *C. macrorhizon* is likely to be limited as leaves are reduced to scales and although the plastome of this species is mostly intact, only the inflorescence axis retains green pigmentation.

Similarly, *Exochaenium oliganthum* (Gilg) Kissling is the sole non-photosynthetic member of its genus and has a photosynthetic sister species, *E. perparvum* (Sileshi) Kissling (Kissling, 2007). The plastome of *E. oliganthum* is also large (151,797 nt) and gene rich, suggesting little degradation. This species retains four *ndh* genes (i.e., *ndhB, ndhC, ndhG, ndhI*) but four photosynthesis-related genes have been pseudogenised (i.e., *psbA, psbB, ccsA* and hypothetical RF3 (*ycf3*) (Darby, 2015). Finally, the genus *Corallorhiza* offers a unique snapshot into how the transition to mycoheterotrophy may impact plastome evolution. There have been at least two independent transitions to holomycoheterotrophy in *Corallorhiza* and based on morphological and plastome sequence differences these species appear to be at different stages of the process. Plastomes in *Corallorhiza* range from 137,505-151,506 nt in length with those of non-green species typically shorter; although this difference was not statistically significant. However, these species do have significantly fewer putatively functional genes (Barrett et al., 2014). The *ndh* genes have been disrupted or lost in all the *Corallorhiza* plastomes characterised to date with additional losses commonly involving photosynthesis genes (e.g., chloroplast envelope membrane protein gene (*cemA*), photosystem I subunit VIII gene (*psaI*), and *psbM*) in the non-green species. Although such gene losses would appear to be consistent with the loss of photosynthesis, some of the genes lost from the plastomes of the non-green species have also been disrupted in at least one of the green and putatively photosynthetic *Corallorhiza* species.

These three groups share broad similarities in terms of size and gene loss patterns with *D. australis*. For example, the plastomes of the non-photosynthetic taxa have lost most of their *ndh* genes. However, in the three orchid groups, close photosynthetic relatives have also lost these genes. Loss of the *ndh* genes is assumed to explain smaller plastome sizes compared to other photosynthetic relatives for *C. macrorhizon* and *C. lancifolium*, as well as *D. australis* and *K. nakaiana*. In these four groups, one (e.g., *ccsA in D. australis*) to several (e.g., *E. oliganthum*) photosynthesis-related genes have also been lost. Some of these losses are shared with other non-photosynthetic species (e.g., loss of *ccsA* from *D. australis* and

E. oliganthum) while others are shared with photosynthetic species (e.g., loss of *ycf15* from the non-photosynthetic *C. macrorhizon* and photosynthetic *G. fumata*).

Indeed, most of the genes identified as disrupted in *Corallorhiza* were also found to be disrupted in at least one of the photosynthetic Goodyerinae examined in the present study. That photosynthesis is clearly still possible in these Goodyerinae implies that, by itself, the loss of individual plastome-encoded photosynthesis-related genes is unlikely to trigger loss of photosynthesis and a transition to holomycoheterotrophy. *Danhachia australis* has acquired many of the morphological features associated with mycoheterotrophy. However, that its plastome remains broadly similar to that of the photosynthetic *K. nakaiana* is a clear indication that it is yet to undergo extensive plastome reduction. This same pattern is also reported for *C. macrorhizon* (Kim et al., 2018) and suggests that, at least in these species, plastome degradation follows, rather than precedes, the transition to heterotrophy.

The Barrett and Davis (2012) model suggests that gene loss in heterotrophic plant plastomes reflects functional relationships between different gene classes. In *D. australis* gene losses are confined to *ndh* and photosynthesis-related genes and in most cases losses from the plastomes of the heterotrophic *Corallorhiza*, *Cymbidium*, and *Exochaenium* species also involve these gene classes. Although *ndh* gene loss is suggested to be the first step in the degradition of heterotrophic plant plastomes, loss of these genes has also been reported from apparently photosynthetic taxa (Kim & Chase, 2017); for example, *ndh* gene loss from the photosynthetic orchid, *K. nakaiana*, is comparable to that of *D. australis* (Table 3.7). Unlike the others, including *Danhatchia* and the available Goodyerinae, the plastid-encoded polymerase (*rpo*) genes have also been disrupted in *Corallorhiza*. However, this observation remains consistent with the Barrett and Davis (2012) model if we assume that *Corallorhiza* is further along in terms of plastome degradation; the *rpo* genes are predicted to be lost following the *ndh* and photosynthesis-related genes.

In terms of gene losses, *C. macrorhizon* and *D. australis* could be thought of as being at a similar stage of plastome degradation. These two species have the least reduced plastomes of any mycoheterotroph investigated to date. Both have lost just a single photosynthesis-related gene in addition to the *ndh* gene class. In *E. oliganthum* several photosynthesis-related genes have been disrupted, suggesting that this species is further along the degradation pathway. Finally, that the *rpo* genes are also disrupted in non-photosynthetic *Corallorhiza* suggests they are further degraded.

Although generally consistent with the Barrett and Davis (2012) model there are exceptions. For example, the ATP synthase CF0 B subunit (*atpF*) is a pseudogene in the photosynthetic *G. procera* despite the model predicting that the *atp* genes would be lost only after the *rpo* genes. Similarly, the housekeeping gene *ycf1* is full lenth in *D. australis* but truncated in several photosynthetic Goodyerinae.

The Barrett and Davis (2012) model predicts stepwise gene loss. The overall pattern reflects a domino effect whereby the functional or physical loss of genes from a given functional class relaxes selection upon, and ultimately triggers the loss of, subsequent gene classes. For example, the functional loss of genes from the photosynthetic gene class is predicted to lead to reduced selective pressure for the maintenance of the *rpo* genes. The *rpo* genes encode the enzyme primarily responsible for transcribing genes within the photosynthetic gene class, while a nuclear encoded RNA-polyermase is responsible for transcription of genes in classes that are lost later.

One prediction of this model is that disruption of the *ndh* gene class should result in relaxed selection, and therefore increased variability, for photosynthesis class genes. We might therefore expect a gradient in gene variability across gene classes. The *ndh* gene class would be the most variable, followed by the photosynthesis gene class, with gene classes affected later in the model being less variable again.

In the Goodyerinae, several genera have experienced *ndh* gene loss (e.g., *Danhatchia*, *Goodyera*, *Kuhlhasseltia*) implying that in this lineage this gene class should exhibit greater variability than the photosynthesis and non-photosynthesis gene classes. However, the present results are not consistent with this expectation. Indeed, statistically significant differences were only found at the level of amino acid identity for two comparisons involving the photosynthetic gene class. This pattern is not well explained by the Barret and Davis model (2012) and instead suggests that the photosynthesis genes are less tolerant of amino acid substitutions than the other classes. Other investigations have also reported that genes in the photosynthesis gene class may be more highly conserved than those in other gene classes, for both photosynthetic and hemiparasitic plants (e.g., McNeal et al., 2007; Nazareno et al., 2015).

Comparative analyses of gene variation suggest that gene losses in *D. australis* are not a continuation of processes typical of photosynthetic genomes.

This is despite the similarities between the plastomes of *D. australis* and the photosynthetic relative *K. nakaiana*. Instead, combined with the observation that other morphologically typical mycoheterotrophs (e.g., *Corallorhiza*) also retain largely intact plastomes, these results imply that plastome degradation begins only after the transition to mycoheterotrophy. If so, given that the *ndh* genes have also been lost from some photosynthetic plant species (e.g., Blazier et al., 2016; Kim & Chase, 2017; Silva et al., 2018), loss of these genes might be better veiwed as predisposing the plastome to further gene loss, rather than as the first step in plastome degradation for heterotrophic plants.

Transitions to holomycoheterotrophy are suggested to have occurred relatively recently in each of *Corallorhiza*, *Cymbidium*, and *Exochaenium*. For *Exochaenium* this is supported by divergence time estimates that imply this genus diverged from its sister lineage ~29 million years ago (mya) and that the subclade containing *E. oliganthum* diverged just ~9 mya (Kissling, 2007). Since all other members of this genus are photosynthetic, *E. oliganthum* probably transitioned to holomycoheterotrophy less than 9 mya. Currently, little is known about the rate at which plastome genes are disrupted in non-photosynthetic plant species. This may vary by lineage or depending on the extent of degradation. However, based on the apparent similarities between the evolution of the *D. australis* and *Exochaenium* plastomes it seems reasonable to assume that the transition to holomycoheterotrophy has also occurred relatively recently in *D. australis*. Addtionally, the GC% of *D. australis* remains very similar to that their photosynthetic relatives. Barrett et al. (2014) suggest that large scale changes in plastome GC% may take longer to manifest in plastome degradation. This also supports the idea that *D. australis* is a recent transition.

3.4.4 A question of origins

The present phylogenetic analyses strongly support the current placement of *D. australis* within the subtribe Goodyerinae (Figure 3.2). However, sampling of Goodyerinae remains insufficient to confidently identify the closest photosynthetic relatives of *D. australis*; the current analyses contain just six of the approximately 450 members of the subtribe (Juswara, 2010). These analyses also highlight an interesting biogeographical issue. The subtribe Goodyerinae is poorly represented in Australia and New Zealand. Indeed, *Danhatchia* is the sole representive of

Goodyerinae in New Zealand. Two other subtribes, Spiranthinae (two species) and Pterostylidinae (upwards of 20 species), are represented in New Zealand. However, the combination of phylogenetic analyses (Figure 3.2) and morphological considerations (e.g., different floral structure in *Pterostylis* R. Br.) suggest neither of these groups are closely related to *D. australis.* These observations make it unlikely that *D. australis* arose in New Zealand, but in this case where and from what did it originate?

One possibility is that *Danhatchia* arose in Australia or New Caledonia. Goodyerinae is represented in both areas, although neither is a centre of diversity. Moreover, most of the taxa recorded from Australia appear to belong or be closely related to the *Goodyera* alliance. In contrast, the taxa that *Danhatchia* clusters most closely with in the present phylogenetic analyses – *L. discolor* and *K. nakaiana* – are Asian taxa outside of the *Goodyera* alliance and are not known to occur in Australia. That said, the recent suggestion that Australian specimens of *Danhatchia* may represent as many as two distinct species (Jones & Clements, 2018) does support an Australian, rather than a New Zealand, origin for this species.

Sequencing of Australian representatives of *Danhatchia* and other Australian Goodyerinae will be an important step towards better understanding the origins of *D. australis*. If *D. australis* were shown to be recently derived from an Australian ancestor this would imply arrival in New Zealand by long distance dispersal. Clearly the diversity and origins of *Danhatchia* require further investigation.

3.5 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403-410.

Banks, D. (2012). *Danhatchia australis* in Australia. *Australian Orchid Review, 77*, 48-49.

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., & Santos, C. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution, 31*, 3095-3112.

Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M. L., & Ruhlman, T. A. (2016). Variable presence of the inverted repeat and plastome stability in *Erodium*. *Annals of Botany, 117*, 1209-1220.

Campbell, E. O. (1972). The morphology of the fungal association of *Corybas cryptanthus*. *Journal of the Royal Society of New Zealand, 2*, 43-47.

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics, 11*, 485.

Darby, H. (2015). *Plastid genome evolution in partially and fully mycoheterotrophic eudicots*. (Text), Retrieved from https://open.library.ubc.ca/collections/24/items/1.0223133.

de Lange, P. J. (2019). *Danhatchia australis* a Fact Sheet (content continuously updated). Retrieved 11 January 2019 from http://www.nzpcn.org.nz/flora_details.aspx?ID=1374.

de Lange, P. J., Rolfe, J. R., Barkla, J. W., Courtney, S. P., Champion, P. D., Perrie, L. R., Beadel, S. M., Ford, K. A., Breitwieser, I., Schonberger, I., Hindmarsh-Walls, R., Heenan, P. B., & Ladley, K. (2017). Conservation status of New Zealand indigenous vascular plants. In *New Zealand Threat Classification Series* (Vol. 22, pp. 22). Wellington: Department of Conservation.

Garay, L. A., & Christensen, E. A. (1995). *Danhatchia* a new genus for *Yoania australis*. *Orchadian, 11*, 469-471.

Garnock-Jones, P. J. (2014). Evidence-based review of the taxonomic status of New Zealand's endemic seed plant genera. *New Zealand Journal of Botany, 52*, 163-212.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology, 59*, 307-321.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology, 52*, 696-704.

Hatch, E. D. (1963). Notes on New Zealand orchids II. *Transactions of the Royal Society of New Zealand, Botany* 185-188.

Jones, D. L., & Clements, M. A. (2018). *Danhatchia novaehollandiae* (Orchidaceae: Goodyerinae), a new species from south-eastern Australia. *Australian Orchid Review, 84*, 56-58.

Juswara, L. S. (2010). *Phylogenetic analyses of subtribe goodyerinae and revision of Goodyera section Goodyera (Orchidaceae) from Indonesia, and fungal association of Goodyera section Goodyera.* The Ohio State University, Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu1275490522

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*, 1647-1649.

Kim, H. T., & Chase, M. W. (2017). Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). *PloS ONE, 12*, e0187318-e0187318.

Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., & Kim, J.-H. (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PloS ONE, 10*, e0142215.

Kim, H. T., Shin, C. H., Sun, H., & Kim, J. H. (2018). Sequencing of the plastome in the leafless green mycoheterotroph *Cymbidium macrorhizon* helps us to understand an early stage of fully mycoheterotrophic plastome structure. *Plant Systematics and Evolution, 304*, 245-258.

Kissling, J. (2007). *Phylogenetics of tribe Exaceae (Gentianaceae) based on molecular, morphological and karyological data, with special emphasis on the genus Sebaea.* (Doctor of Philosophy), University of Neuchâtel,

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H.,, Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*, 2.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*, 1754-1760.

Lin, C. S., Chen, J. J. W., Huang, Y. T., Chan, M. T., Daniell, H., Chang, W. J., Hsu, C. T., Liao, D. C., Wu, F. H., Lin, S. Y., Liao, C. F., Deyholos, M. K., Wong, G. K. S., Albert, V. A., Chou, M. L., Chen, C. Y., & Shih, M. C. (2015). The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Scientific Reports, 5*, 9040.

Lohse, M., Drechsel, O., Kahlau, S., & Bock, R. (2013). OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research, 41*, 575-581.

McNeal, J. R., Kuehl, J. V., Boore, J. L., & De Pamphilis, C. W. (2007). Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC plant biology, 7*

Menezes, A. P. A., Resende-Moreira, L. C., Buzatti, R. S. O., Nazareno, A. G., Carlsen, M., Lobo, F. P., Kalapothakis, E., & Lovato, M. B. (2018). Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Scientific Reports, 8*, 2210.

Merckx, V. (2012). *Mycoheterotrophy: The Biology of Plants Living on Fungi*. New York: Springer-Verlag.

Merckx, V., Bidartondo, M. I., & Hynson, N. A. (2009). Myco-heterotrophy: when fungi host plants. *Annals of Botany, 104*, 1255-1261.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., & Marshall, D. (2010). Tablet—next generation sequence assembly visualization. *Bioinformatics, 26*, 401-402.

Moore, L. B., & Edgar, E. (1970). *Indigenous Tracheophyta* (Vol. II). Wellington, New Zealand: Government Printer.

Nazareno, A. G., Carlsen, M., & Lohmann, L. G. (2015). Complete chloroplast genome of *Tanaecium tetragonolobum*: The first Bignoniaceae plastome. *PloS ONE, 10*, e0129930.

Park, I., Yang, S., Kim, W. J., Noh, P., Lee, H. O., & Moon, B. C. (2018). The complete chloroplast genomes of six *Ipomoea* species and indel marker development for the discrimination of authentic *Pharbitidis* semen (Seeds of *I. nil* or *I. purpurea*). *9*

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics, 28*, 1420-1428.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology Evolution, 25*, 1253-1256.

Pridgeon, A. M., Cribb, P. J., Rasmussen, F. N., & Chase, M. W. (1999). In *Genera Orchidacearum: Orchidoideae (Part Two) Vanilloideae* (Vol. 3, pp. 358). Oxford: Oxford University Press.

Schelkunov, M. I., Shtratnikova, V. Y., Nuraliev, M. S., Selosse, M. A., Penin, A. A., & Logacheva, M. D. (2015). Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution, 7*, 1179-1191.

Schwarz, G. (1978). Estimating the dimenions of a model. *Annals of Statistics, 6*, 461-464.

Silva, S. R., Michael, T. P., Meer, E. J., Pinheiro, D. G., Varani, A. M., & Miranda, V. F. O. (2018). Comparative genomic analysis of *Genlisea* (corkscrew plants—Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes. *PloS ONE, 13*, e0190321.

Snedecor, G. W., & Cochrane, W. G. (1989). *Statistical Methods* (Eighth ed.). USA: Iowa State University Press.

Steenbeeke, G. (2012). *Danhatchia australis* - not just a New Zealand species. *Orchadian, 17*

Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany, 99*, 349-364.

Suetsugu, K., Ohta, T., & Tayasu, I. (2018). Partial mycoheterotrophy in the leafless orchid *Cymbidium macrorhizon*. *American Journal of Botany, 105*, 1595-1600.

Twyford, A. D., & Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular Ecology Resources, 17*, 858-868.

Zhu, S., Niu, Z., Yan, W., Xue, Q., & Ding, X. (2016). The complete chloroplast genome sequence of *Anoectochilus emeiensis*. *Mitochondrial DNA A DNA Mapp Seq Anal, 27*, 3565-3566.

Chapter 4

Conclusions

This thesis reports complete plastome sequences for two holomycoheterotrophic orchids endemic to New Zealand, *Corybas cryptanthus* Hatch and *Danhatchia australis* (Hatch) Garay & Christenson, as well as the plastomes of eleven close photosynthetic relatives of *C. cryptanthus*. Comparisons of these sequences to those already available for related plant species provide insights into plastome evolution in these New Zealand mycoheterotrophs, and potentially, in mycoheterotrophic plants more generally.

4.1 Main findings

4.1.1 The plastomes of C. cryptanthus and D. australis differ in the extent of degradation

Plastome degradation in the form of reduced genome size, structural rearrangements, and functional gene loss is characteristic of most heterotrophic plant species (Graham et al., 2017). The Barrett and Davis (2012) model describes a stepwise sequence of gene loss in heterotrophic plastomes that reflects functional considerations. For example, the NADH dehydrogenase subunit (*ndh*) genes, which are also absent from some photosynthetic plastomes, are lost first, while at least some genes from the final gene class to be lost, the so-called housekeeping genes, are retained even in the smallest known heterotrophic plastomes (Barrett & Davis, 2012).

Despite broad similarities in morphology (e.g., a lack of green pigmentation, inflorescence the only aboveground portion) the plastomes of *C. cryptanthus* and *D. australis* differ markedly in terms of genome size, extent of structural rearrangement, and degree of functional gene loss compared to photosynthetic relatives. Due to its taxonomic isolation, the plastome of *D. australis* was expected to exhibit greater reduction. However, this was not the case. Instead, the plastome of *C. cryptanthus* is highly modified, being about half the size of its photosynthetic congeners as well as having lost an entire inverted repeat (IR) copy and numerous genes relative to other *Corybas* Salisb. species. In contrast, the plastome of

97

D. australis is very similar in size, structure and gene content to photosynthetic genera of the Goodyerinae. Clearly, neither taxonomic nor phylogenetic isolation are not good predictors of plastome reduction.

Both the *C. cryptanthus* and *D. australis* plastomes are broadly consistent with the Barrett and Davis (2012) model, albeit representing different stages. The *ndh* gene class, the first to be affected in the model, has been entirely (i.e., *C. cryptanthus*) or almost entirely (i.e., *D. australis*) lost from these plastomes. However, the implications of this observation may differ between these species. While the *ndh* genes were absent from all characterised *Corybas* plastomes, all but one available Goodyerinae plastome retains at least ten of the 11 *ndh* genes. Specifically, Kim et al. (2015) and Niu et al. (2017) have suggested that the physical loss of the *ndh* genes may destabilise the plastome. If so, the loss of these genes in *Corybas* as a whole may have predisposed the ancestor of *C. cryptanthus* to rapid plastome evolution. In contrast, the photosynthetic ancestor of *Danhatchia* Garay & Christenson seems likely to have had a full set of *ndh* genes, perhaps slowing the degradation process.

The plastome of *D. australis* has lost a single gene from the photosynthesis gene class, which according to the Barrett and Davis (2012) model is the next to be lost following the *ndh* genes. In contrast, most of this class and all those from the next two gene classes (i.e., the adenosine triphosphate (ATP) synthase subunit (*atp*) and plastid-encoded RNA polymerase subunit (*rpo*) genes) are absent from the *C. cryptanthus* plastome. Although broadly consistent with the Barrett and Davis (2012) model, neither plastome adheres strictly to the stepwise pattern of gene losses predicted by it. In both cases, losses from a later gene class have occurred before the complete loss of genes from an earlier class. This suggests that selective pressures may be relaxed on subsequent gene classes before complete loss of the earlier class or classes. This observation is consistent with the findings of Barrett et al. (2014) and Graham et al. (2017) that suggest there is broad overlap between the photosynthesis, *rpo*, and *atp* gene classes in terms of the relative timing of gene loss.

The differences between the *C. cryptanthus* and *D. australis* plastomes may reflect the timing of transition to mycoheterotrophy. The plastome of *D. australis* is similar to that of *Cymbidium macrorhizon* Lindl. that is thought to have transitioned recently (Kim et al., 2018). In contrast, the plastome of *C. cryptanthus* is more

similar to that of *Rhizanthella gardneri* R. S. Rogers (Delannoy et al., 2011) that is thought to have been non-photosynthetic for a longer period of time.

4.1.2 Plastomes of the two C. cryptanthus accessions differ at the sequence level

The plastome sequences of the two *C. cryptanthus* accessions were broadly similar. Both have the same gene content and all but three protein-coding genes were identical in length and nucleotide sequence. They also share loss of one IR copy; this loss is therefore likely to have occurred in a common ancestor of the two populations. These plastomes differed in overall length due to indel differences and in DNA sequence due to nucleotide substitutions. Similar differences have been reported between accessions of *Monotropa hypopitys* L. (Gruzdev et al., 2016; Ravin et al., 2016) and *Epipogium roseum* (D. Don). Lindl (Schelkunov et al., 2015).

The numbers of nucleotide substitutions and indels differentiating the two *C. cryptanthus* plastomes are similar to those observed between closely related photosynthetic *Corybas* (e.g., *C. obscurus* Lehnebach and *C.* "Remutaka"). In contrast, there are substantially more nucleotide substitutions and indels between the IR analogues of *C. cryptanthus* and the IR regions of photosynthetic *Corybas*, than there are between the IR regions of photosynthetic *Corybas* species. This suggests an accelerated rate of change in *C. cryptanthus*, perhaps resulting from loss of the second IR copy (c.f.,Wolfe et al., 1987; Xu et al., 2015; Zhu et al., 2016). This observation is consistent with the suggestion that the IR regions act to stabilise the plastome (Palmer & Thompson, 1981; Palmer & Thompson, 1982).

The analyses of Lyon (2015) suggest that *C. cryptanthus* diverged from its sister lineage ~9 million years ago (mya). In contrast, the photosynthetic New Zealand species have arisen over the last 4 million years. If the *C. cryptanthus* plastome is evolving more quickly than those of other *Corybas*, it is reasonable to assume that the two examined accessions of *C. cryptanthus* diverged comparatively recently. Further sampling will be needed to establish the overall genetic diversity of *C. cryptanthus*. However, the present results are consistent with the *C. cryptanthus* lineage persisting in New Zealand for an extended period without diversifying or, alternatively, with early diverging lineages having been lost to extinction.

4.1.3 The plastomes of photosynthetic Corybas are structurally unusual

The plastomes of photosynthetic *Corybas* have the quadripartite structure typical of photosynthetic land plants. However, in *Corybas* the IR regions are larger and the small single copy (SSC) regions smaller than in many other photosynthetic land plants. Indeed, the SSC region in photosynthetic *Corybas* is smaller than any previously reported for a photosynthetic angiosperm (Park et al., 2018). Almost all the genes typically found in the SSC regions of photosynthetic orchids (e.g., *Microtis unifolia* G.Forst and *Chiloglottis cornuta* Hook.f.) are incorporated into the IRs in these species.

In the available plastomes for Goodyerinae, including *D. australis*, the boundaries between the SSC and IR regions were stable, In contrast, these boundaries varied across the sequenced Diurideae. Complete coding sequences for 5-6 genes fall within the SSC regions of *Ch. cornuta* and *M. unifolia*, but at most two complete coding sequences fall within the SSC regions of photosynthetic *Corybas*. The substantial expansion of the IRs and dramatic reduction of the SSC region are therefore linked, as they are in *Lamprocapnos spectabilis* L. (Papaveraceae; Park et al., 2018) and *Pedicularis ishiodoyana* Koids. & Ohwi (Orobanchaceae; Cho et al., 2018). Destabilisation of the IR has been linked to the physical loss of the *ndh* genes and particularly of *ndhF* (Kim et al., 2015; Niu et al., 2017). However, the evolutionary implications of reduced SSC regions in *Corybas* are uncertain. It is not clear whether this structural change has impacted upon the pattern or rate of change in the *C. cryptanthus* plastome, although this certainly seems possible.

4.1.4 Morphological and trophic changes precede plastome degradation

A key evolutionary question for studies of mycoheterotrophy is whether the transition to this lifestyle is driven by changes to the plastome or whether such changes follow the trophic shift. Consistent with Kim et al. (2018) several observations on the present analyses support the hypothesis that plastome change follows the lifestyle shift.

Perhaps the most obvious is that the *D. australis* plastome remains little changed. Aside from the *ndh* genes, only one gene otherwise present in photosynthetic Goodyerinae has been functionally lost from the *D. australis* plastome (i.e., the cytochrome biogenesis c gene, *ccsA*). It is possible that the

disruption of *ccsA* is responsible for the loss of photosynthesis. However, this seems unlikely given that *D*. *australis* has many of the morphological features typical of holomycoheterotrophs (e.g., leaflessness and subterranean lifestyle). For *ccsA* disruption to have initiated the transition to mycoheterotrophy would require these morphological changes to have occurred without further disruption to the plastome. Similarly low numbers of gene losses have been reported for mycoheterotrophic species such as *C*. *macrorhizon* (Kim et al., 2018) and *Corallorhiza mertensiana* Bong (Barrett et al., 2014) suggesting that in general plastome degradation follows other changes (e.g., morphological and physiological).

The plastome of the albino *C*. "Waitarere white" is also little different from those of photosynthetic congeners. Yet, while these plants lack pigmentation (c.f., being non-green), their aboveground morphology is otherwise apparently normal. Albinism in these individuals is therefore unlikely to be the result of plastome disruption and may instead be explained by changes in the nuclear genome. Additionally, it is unclear whether *C*. "Waitarere white" individuals persist long-term and therefore whether these plants represent a step along the pathway towards holomycoheterotrophy. Similar questions arise for albino representatives of other, otherwise photosynthetic Orchidaceae. For example, *Cephalanthera longifolia* (L.) Fritsch (Abadie et al., 2006) and *Epipactis helleborine* (L.) Crantz (Suetsugu et al., 2017). Although plastomes are not yet available for these species, albino individuals are known to persist for several years.

Comparisons of plastome gene content for the available Diurideae and Goodyerinae are also consistent with the idea that plastome gene loss follows, rather than initiates, the transition to mycoheterotrophy. Consistent with the Barrett and Davis (2012) model, the *ndh* genes were disrupted in all three non-photosynthetic taxa investigated. However, theses genes are also disrupted in photosynthetic relatives (e.g., all Diurideae investigated and *Kuhlhasseltia nakaiana* (F. Maek.) Ormerod). Similarly, all photosynthetic *Corybas* lack *ycf15*, a photosynthesis class gene in the Barrett and Davis (2012) model, while several photosynthetic Goodyerinae lack photosynthesis and RNA polymerase class genes. Comparative analyses conducted here and elsewhere (Magee et al., 2010; Nevill et al., 2019; Rousseau-Gueutin et al., 2013) suggest that photosynthetic land plants may tolerate the loss of certain plastome-encoded genes without loss of photosynthesis or

wholesale plastome degradation. For example, *Goodyera procera* (Ker Gawl.) Hook and *K. nakaiana* remain photosynthetic despite losing the same number of genes as *D. australis*. Surprisingly, in *G. procera* only one loss involved the *ndh* gene class whereas in *D. australis* only one loss did not involve the *ndh* gene class.

4.2 Future work

The present analyses provide initial insights into plastome evolution in two New Zealand endemic holomycoheterotrophic orchids. These suggest that, rather than causing loss of photosynthesis and driving trophic change, plastome degradation follows the transition to mycoheterotrophy. There are several possible avenues for further work with these native New Zealand plants.

One avenue involves further investigations of DNA sequence diversity. To date, few investigations have examined intraspecific variation in mycoheterotroph plastomes. Although the plastomes of the two characterised *C. cryptanthus* accessions differed, plastomes from additional individuals are needed to assess the significance of these differences. In the case of *Danhatchia*, additional New Zealand accessions as well as material from Australian plants recently recognised as new species (Jones & Clements, 2018) would be needed. Investigating the nuclear genomes of *C. cryptanthus* and *D. australis* will also be important for understanding the transition to mycoheterotrophy in these species. The present analyses suggest that changes to the plastome have followed the transition to mycoheterotrophy. Many chloroplast-associated genes are nuclear-encoded and changes to these could be central to the loss of photosynthesis in mycoheterotrophic plants. One approach could be to compare sequences from functional copies of nuclear-encoded chloroplast-expressed genes to those from *C. cryptanthus* and *D. australis*. However, this approach is limited by a lack of potential reference sequences, particularly from orchids. An alternative might be comparative transcriptomics of photosynthetic and non-photosynthetic species.

It would also be interesting to study the albino *C.* "Waitarere white" in more detail. Albinism has been recorded for other Orchidaceae (Abadie et al., 2006; Julou et al., 2005; Salmia, 1989; Selosse et al., 2004) but this is the first known record of albino individuals in *Corybas*. An initial step would be to identify the taxon or taxa

from which the albino individuals were derived, perhaps using targeted sequencing of chloroplast markers. Direct comparisons of the chloroplast and nuclear genomes or transcriptomes of albino and green individuals of the same species could then be used to examine the origins of the albino forms. Likewise, it would be interesting to track the persistence of *C.* "Waitarere white" individuals over time to assess their evolutionary significance.

Beyond further genetic characterisation, future work should also consider other aspects of the transition to holomycoheterotrophy in *C. cryptanthus* and *D. australis*. In other orchid groups, fungal partners of holomycoheterotrophs differ from those of photosynthetic relatives (Merckx, 2012; Merckx et al., 2009; Ogura-Tsujita et al., 2012). One research line might involve characterising the relationship between the plant and fungal partners for these New Zealand species. The fungal partners of *C. cryptanthus* and *D. australis* are unknown. Whether these species form associations with the same fungal species as photosynthetic relatives is also uncertain. Understanding these relationships may help identify the pathway to holomycoheterotrophy. Comparative analyses to identify morphological and physiological differences, in particular the relative contributions of photosynthesis and the fungal partner between the holomycoheterotrophic and photosynthetic species, would also be useful, as would be similar analyses for the albino *Corybas*.

4.3 References

Abadie, J. C., Püttsepp, Ü., Gebauer, G., Faccio, A., Bonfante, P., & Selosse, M. A. (2006). *Cephalanthera longifolia* (Neottieae, Orchidaceae) is mixotrophic: a comparative study between green and nonphotosynthetic individuals. *Canadian Journal of Botany, 84*, 1462-1477.

Barrett, C. F., & Davis, J. I. (2012). The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany, 99*, 1513-1523.

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., & Santos, C. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution, 31*, 3095-3112.

Cho, W. B., Choi, B. H., Kim, J. H., Lee, D. H., & Lee, J. H. (2018). Complete plastome sequencing reveals an extremely diminished SSC region in hemiparasitic *Pedicularis ishiodoyana* (Orobanchaceae). *Annales Botanici Fennici, 55*, 171-183.

Delannoy, E., Fujii, S., Colas Des Francs-Small, C., Brundrett, M., & Small, I. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution, 28*, 2077-2086.

Graham, S. W., Lam, V. K. Y., & Merckx, V. S. F. T. (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist, 214*, 48-55.

Gruzdev, E., Mardanov, A., Beletsky, A., Kadnikov, V., Kochieva, E., Ravin, N., & Skryabin, K. (2016). Reduction of the chloroplast genome and the loss of photosynthetic pathways in the mycoheterotrophic plant *Monotropa hypopitys*, as revealed by genome and transcriptome sequencing. *Febs Journal, 283*, 341-342.

Jones, D. L., & Clements, M. A. (2018). *Danhatchia novaehollandiae* (Orchidaceae: Goodyerinae), a new species from south-eastern Australia. *Australian Orchid Review, 84*, 56-58.

Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., & Selosse, M. A. (2005). Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytologist, 166*, 639-653.

Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., & Kim, J.-H. (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PloS ONE, 10*, e0142215.

Kim, H. T., Shin, C. H., Sun, H., & Kim, J. H. (2018). Sequencing of the plastome in the leafless green mycoheterotroph *Cymbidium macrorhizon* helps us to understand an early stage of fully mycoheterotrophic plastome structure. *Plant Systematics and Evolution, 304*, 245-258.

Lyon, P. S. (2015). *Molecular systematic, biogeography, and mycorrhizal associations in the Acianthinae (Orchidaceae), with a focus on the genus Corybas.* (PhD Thesis), University of Wisconsin-Madison, USA.

Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Semon, M., Perry, A. S., Stefanovic, S., Milbourne, D., Barth, S., Palmer, J. D., Gray, J. C., Kavanagh, T. A., & Wolfe, K. H. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research, 20*, 1700-1710.

Merckx, V. (2012). *Mycoheterotrophy: The Biology of Plants Living on Fungi.* New York: Springer-Verlag.

Merckx, V., Bidartondo, M. I., & Hynson, N. A. (2009). Myco-heterotrophy: when fungi host plants. *Annals of Botany, 104*, 1255-1261.

Nevill, P. G., Howell, K. A., Cross, A. T., Williams, A. V., Zhong, X., Tonti-Filippini, J., Boykin, L. M., Dixon, K. W., & Small, I. (2019). Plastome-wide rearrangements and gene losses in carnivorous Droseraceae. *Genome Biology and Evolution, 11*, 472-485.

Niu, Z., Pan, J., Zhu, S., Li, L., Xue, Q., Liu, W., & Ding, X. (2017). Comparative analysis of the complete plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids*., 8*, 1713.

Ogura-Tsujita, Y., Yokoyama, J., Miyoshi, K., & Yukawa, T. (2012). Shifts in mycorrhizal fungi during the evolution of autotrophy to mycoheterotrophy in *Cymbidium* (Orchidaceae). *American Journal of Botany, 99*, 1158-1176.

Palmer, J. D., & Thompson, W. F. (1981). Rearrangements in the chloroplast genomes of mung bean and pea. *Proceedings of the National Academy of Sciences of the United States of America, 78*, 5533-5536.

Palmer, J. D., & Thompson, W. F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell, 29*, 537-550.

Park, Am, & Park. (2018). Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion and intraspecific variation. *Scientific Reports, 8*, 13568.

Ravin, N. V., Gruzdev, E. V., Beletsky, A. V., Mazur, A. M., Prokhortchouk, E. B., Filyushin, M. A., Kochieva, E. Z., Kadnikov, V. V., Mardanov, A. V., & Skryabin, K. G. (2016). The loss of photosynthetic pathways in the plastid and nuclear

genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC plant biology, 16*, 153–161.

Rousseau-Gueutin, M., Huang, X., Higginson, E., Ayliffe, M., Day, A., & Timmis, J. N. (2013). Potential functional replacement of the plastidic Acetyl-CoA Carboxylase subunit *(accD)* gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology, 161*, 1918-1929.

Salmia, A. (1989). General morphology and anatomy of chlorophyll-free and green forms of *Epipactis helleborine* (Orchidaceae). *Annales Botanici Fennici, 26*, 95-105.

Schelkunov, M. I., Shtratnikova, V. Y., Nuraliev, M. S., Selosse, M. A., Penin, A. A., & Logacheva, M. D. (2015). Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution, 7*, 1179-1191.

Selosse, M. A., Faccio, A., Scappaticci, G., & Bonfante, P. (2004). Chlorophyllous and achlorophyllous specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) are associated with ectomycorrhizal septomycetes, including truffles. *Microbial Ecology, 47*, 416-426.

Suetsugu, K., Yamato, M., Miura, C., Yamaguchi, K., Takahashi, K., Ida, Y., Shigenobu, S., & Kaminaka, H. (2017). Comparison of green and albino individuals of the partially mycoheterotrophic orchid *Epipactis helleborine* on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Molecular Ecology, 26*, 1652-1669.

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America, 84*, 9054-9058.

Xu, J. H., Liu, Q., Hu, W., Wang, T., Xue, Q., & Messing, J. (2015). Dynamics of chloroplast genomes in green plants. *Genomics, 106*, 221-231.

Zhu, A., Guo, W., Gupta, S., Fan, W., & Mower, J. P. (2016). Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist, 209*, 1747-1756.

Appendix I

**Supplementary Table 1 Publicly available whole plastome sequences for Orchidaceae**

Species	Genbank accession number
Anoectochilus roxburghii	KR779936
Aphyllorchis montana	NC_030703
Apostasia odorata	NC_030722
Bletilla ochracea	NC_029483
Bletilla striata	NC_028422
Calanthe triplicata	NC_024544
Cattleya crispata	NC_026568
Cattleya liliputana	KP202881
Cephalanthera humilis	NC_030706
Cephalanthera longifolia	NC_030704
Corallorhiza bulbosa	NC_025659
Corallorhiza macrantha	NC_025660
Corallorhiza maculata	KM390014
Corallorhiza mertensiana	NC_025661
Corallorhiza odontorhiza	NC_025664
Corallorhiza striata	JX087681
Corallorhiza trifida	NC_025662
Corallorhiza wisteriana	NC_025663
Cymbidium aloifolium	NC_021429
Cymbidium ensifolium	KU179434
Cymbidium faberi	NC_027743
Cymbidium goeringii	NC_028524
Cymbidium kanran	NC_029711
Cymbidium lancifolium	NC_029712
Cymbidium macrorhizon	NC_029713
Cymbidium mannii	NC_021433
Cymbidium sinense	NC_021430
Cymbidium tortisepalum	NC_021431
Cymbidium tracyanum	NC_021432
Cypripedium formosanum	NC_026772
Cypripedium japonicum	NC_027227
Cypripedium macranthos	NC_024421
Dendrobium catenatum	KX507360
Dendrobium chrysotoxum	NC_028549
Dendrobium huoshanense	NC_028430
Dendrobium nobile	NC_029456
Dendrobium officinale	KJ862886
Dendrobium pendulum	NC_029705
Dendrobium strongylanthum	NC_027691
Elleanthus sodiroi	KR260986
Epipactis mairei	NC_030705

|	
Epipactis veratrifolia	NC_030708
Erycina pusilla	NC_018114
Goodyera fumata	NC_026773
Goodyera procera	NC_029363
Goodyera schlechtendaliana	NC_029364
Goodyera velutina	NC_029365
Habenaria pantlingiana	NC_026775
Listera fugongensis	NC_030711
Ludisia discolor	NC_030540
Masdevallia coccinea	KP205432
Masdevallia picturata	NC_026777
Neottia acuminata	NC_030709
Neottia camtschatea	NC_030707
Neottia listeroides	NC_030713
Neottia ovata	NC_030712
Neottia pinetorum	NC_030710
Neuwiedia singapureana	KM244735
Oncidium 'Gower Ramsey'	NC_014056
Oncidium sphacelatum	NC_028148
Paphiopedilum armeniacum	KT388109
Paphiopedilum niveum	NC_026776
Phalaenopsis aphrodite	AY916449
Phalaenopsis equestris	NC_017609
Phalaenopsis 'Tiny Star'	NC_025593
Phragmipedium longifolium	NC_028149
Sobralia aff *bouchei*	NC_028209
Sobralia callosa	NC_028147
Vanilla planifolia	NC_026778
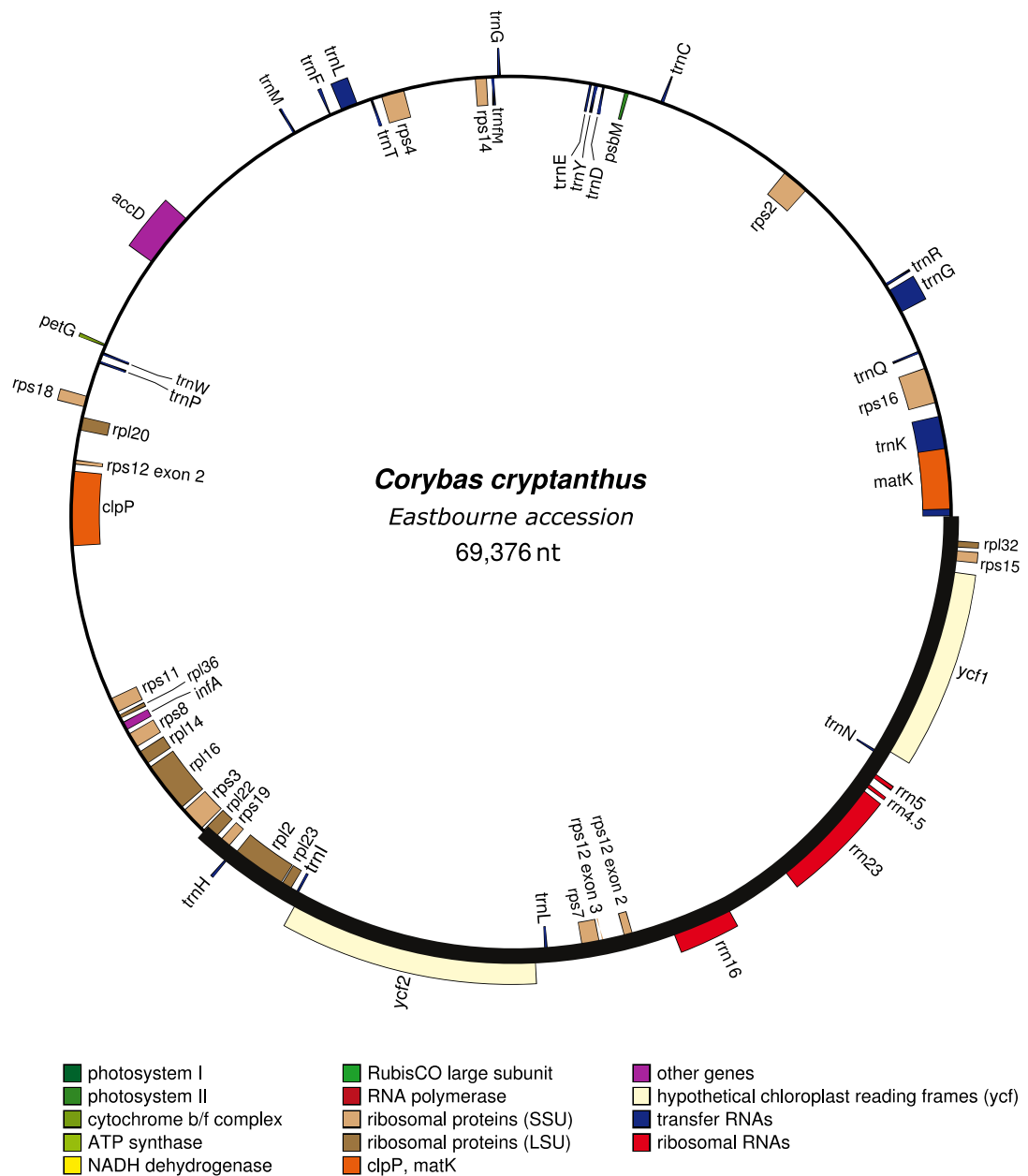
Supplementary Figure 1 The plastome of *Corybas cryptanthus* Eastbourne accession Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of the circle are transcribed in the reverse direction. The region of the plastome equivalent to the IR in photosynthetic *Corybas* (the "IR analogue") is emboldened. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013)
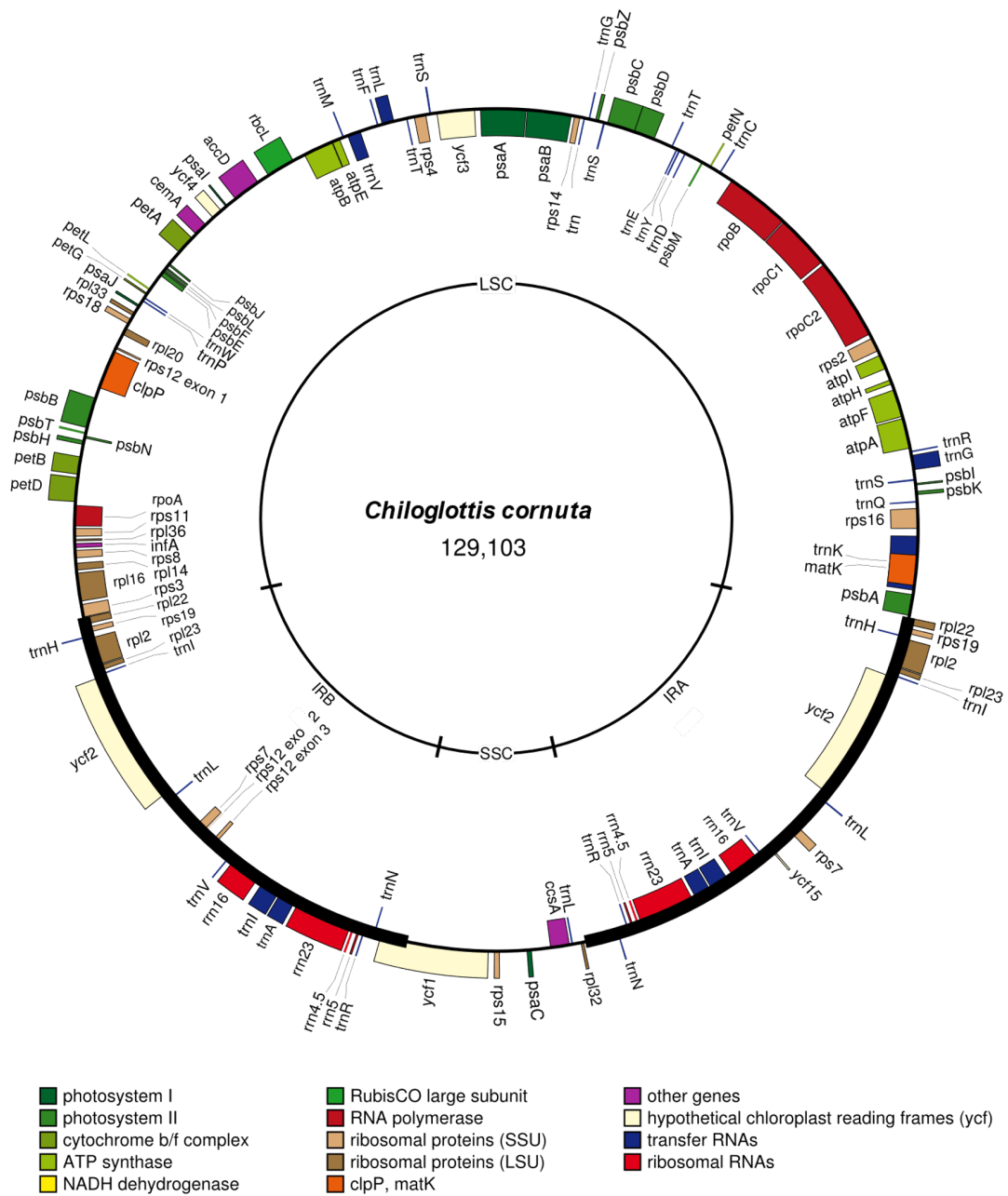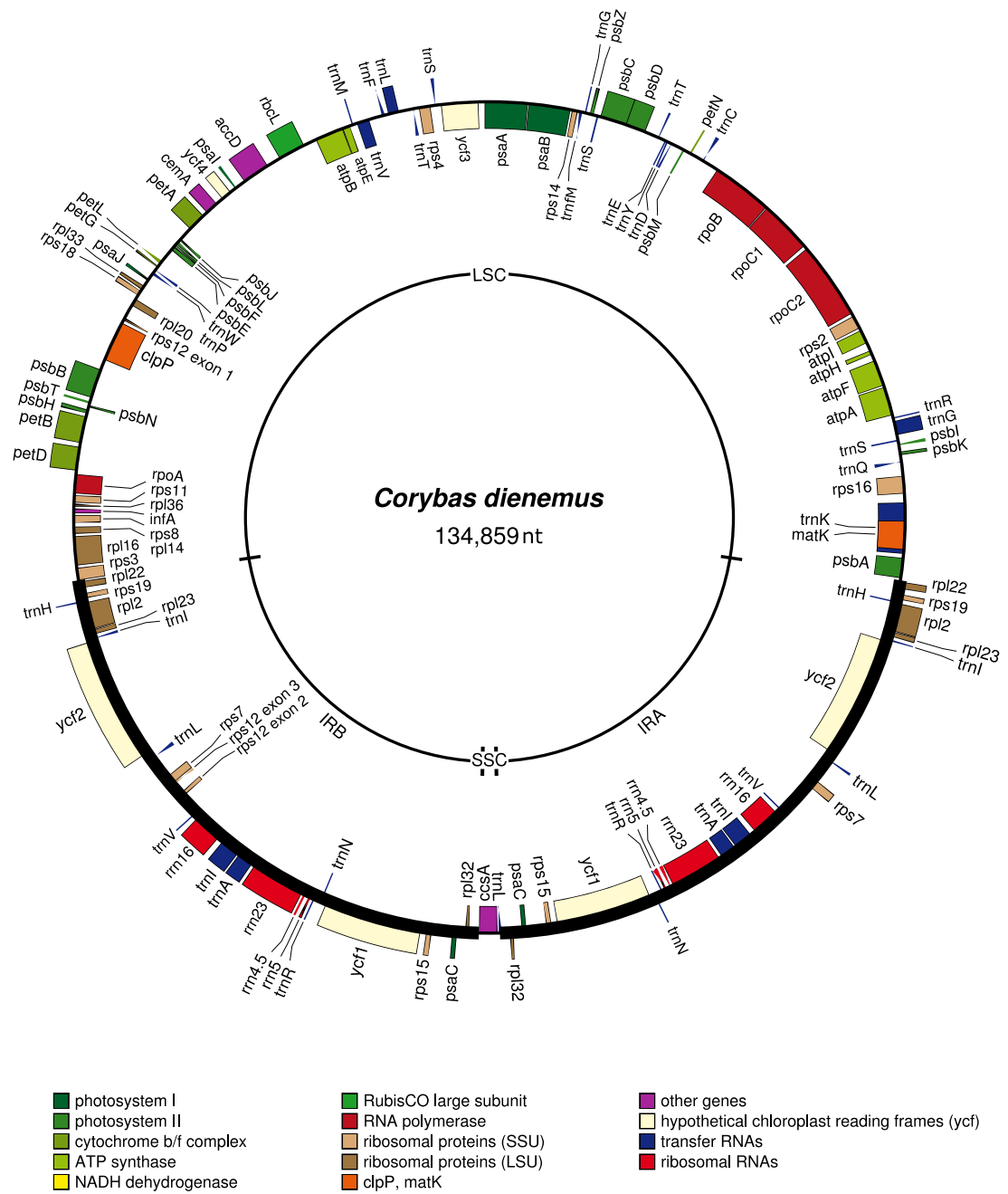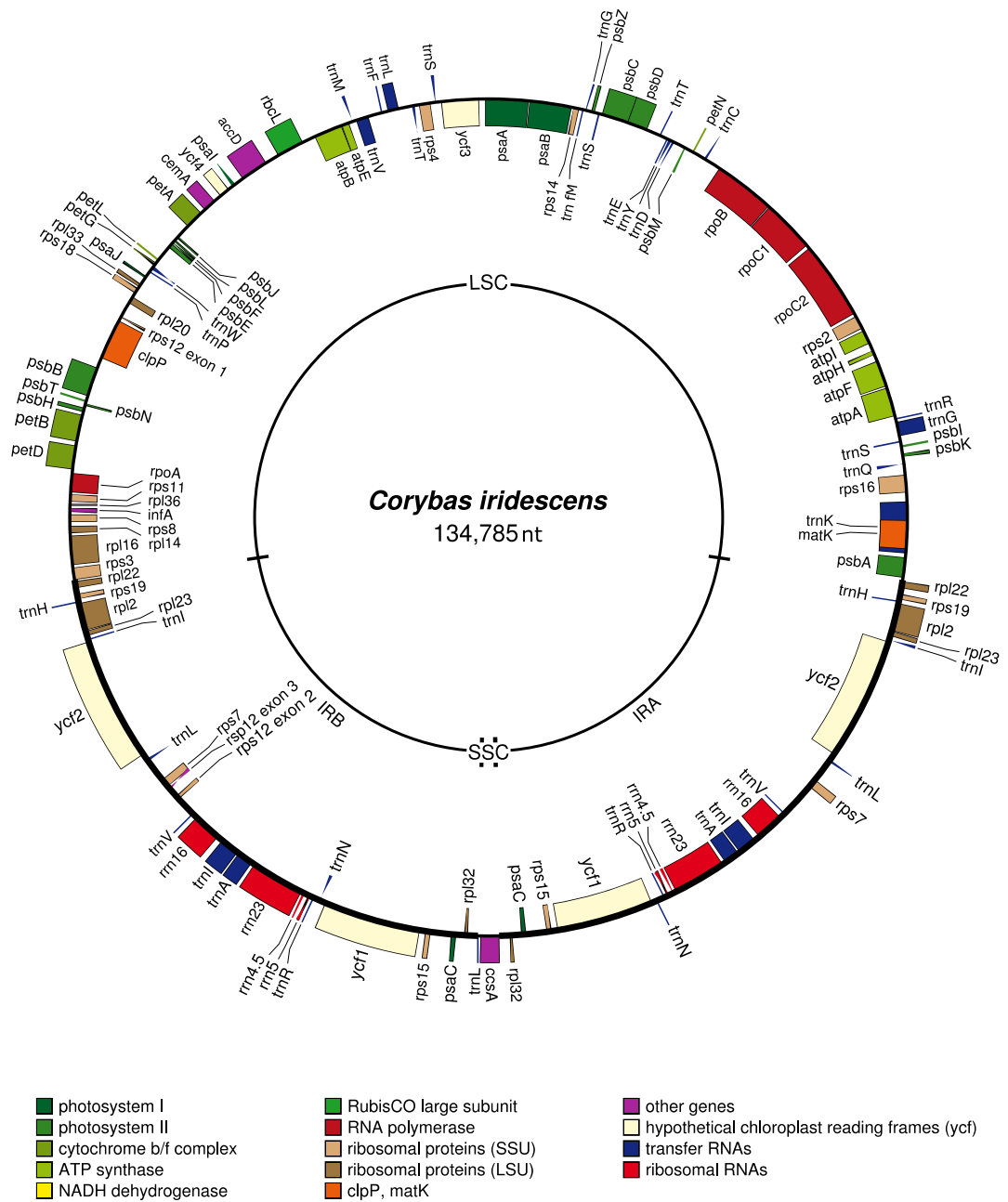
Supplementary Figure 2 The plastome of *Chiloglottis cornuta* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013)
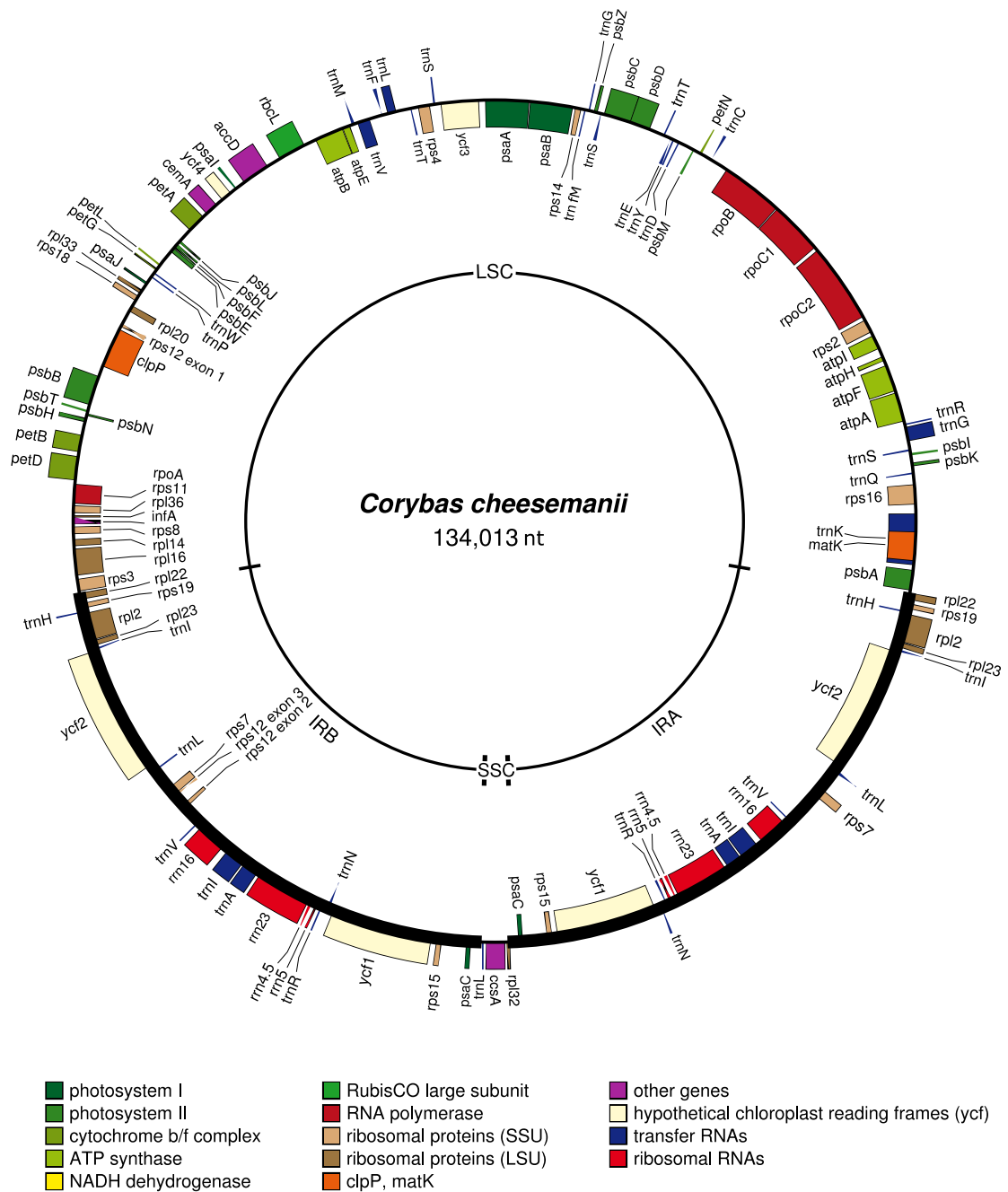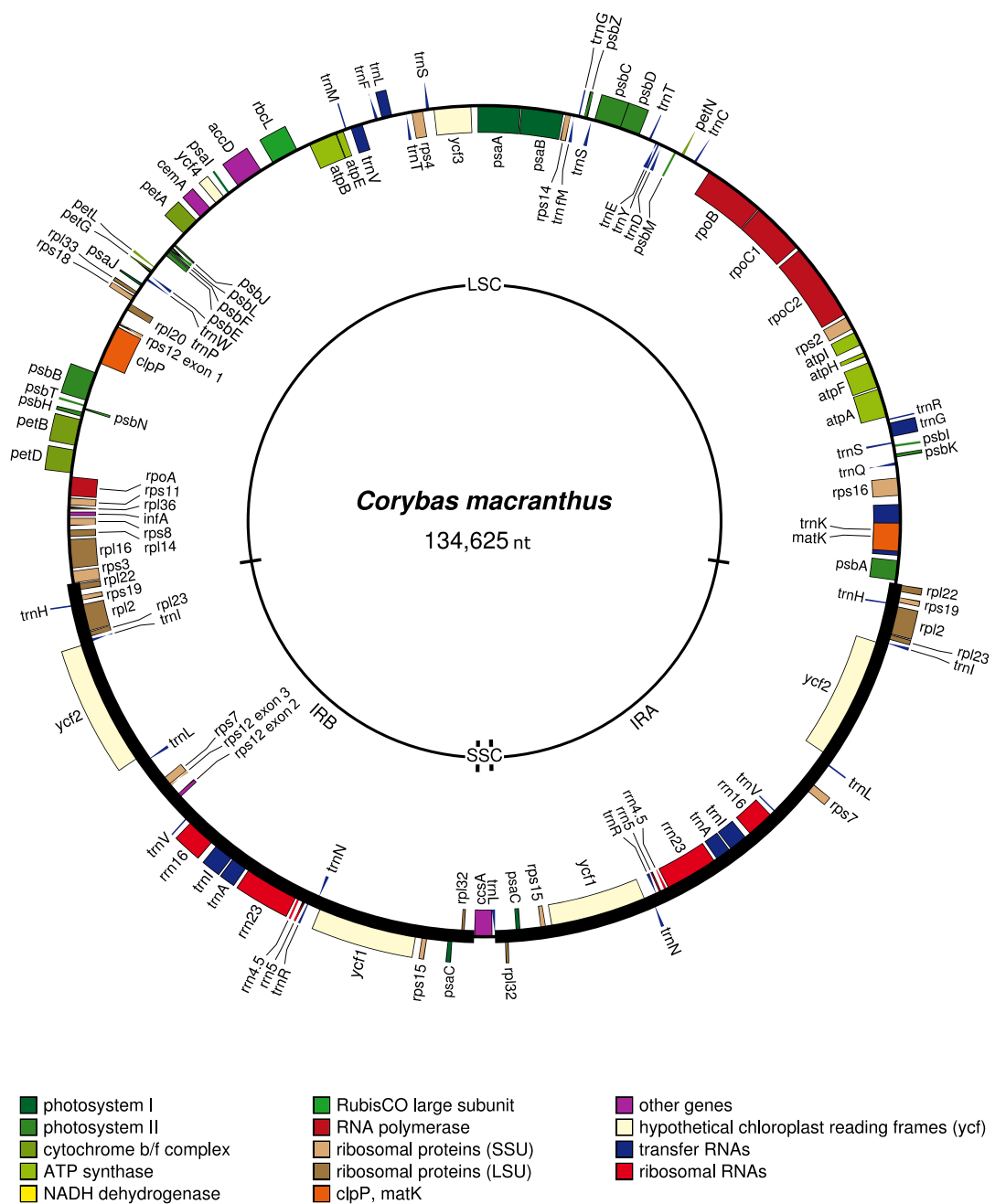
Supplementary Figure 3 The plastome of *Corybas cheesemanii* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013)
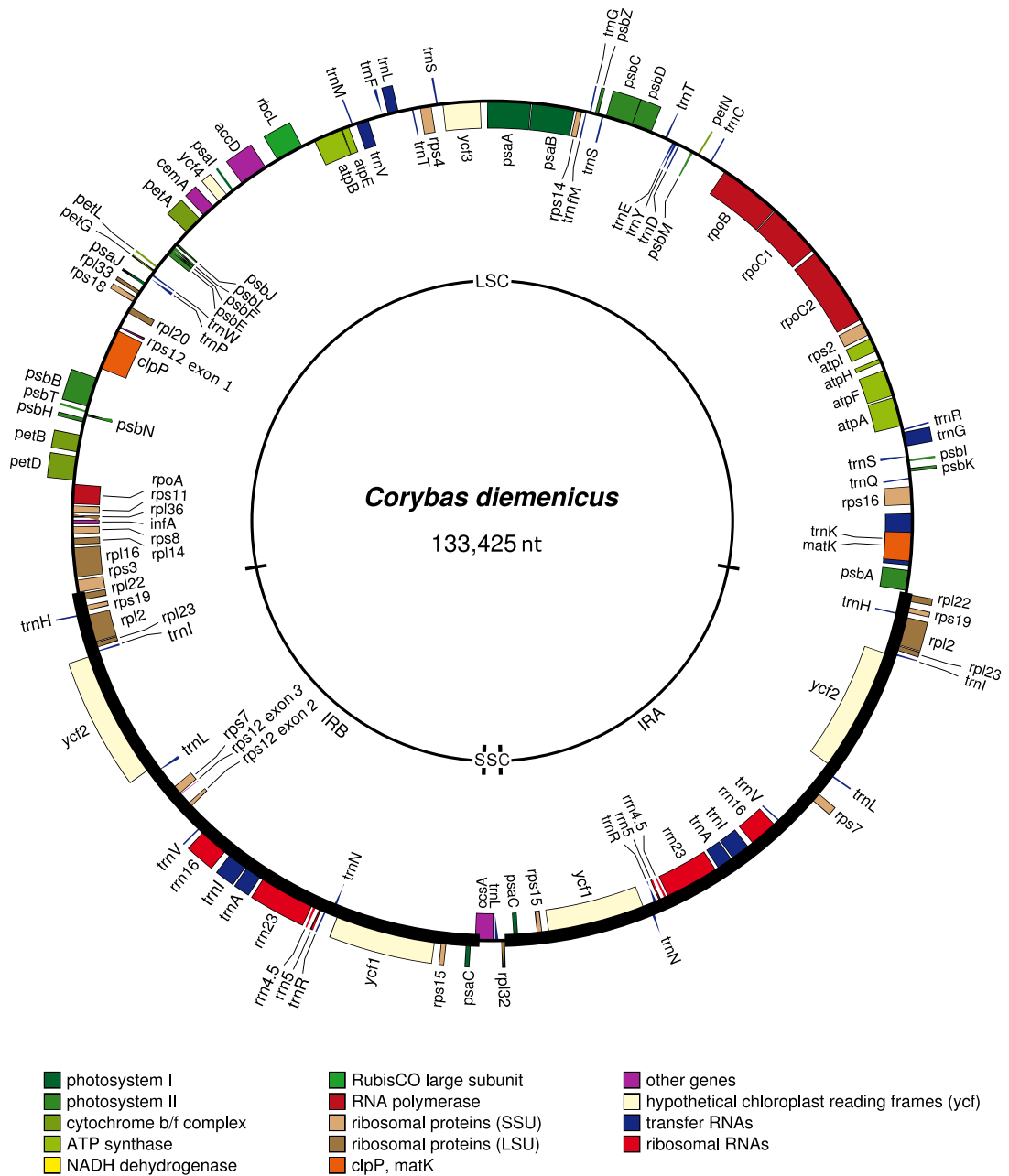
Supplementary Figure 4 The plastome of *Corybas diemenicus* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).
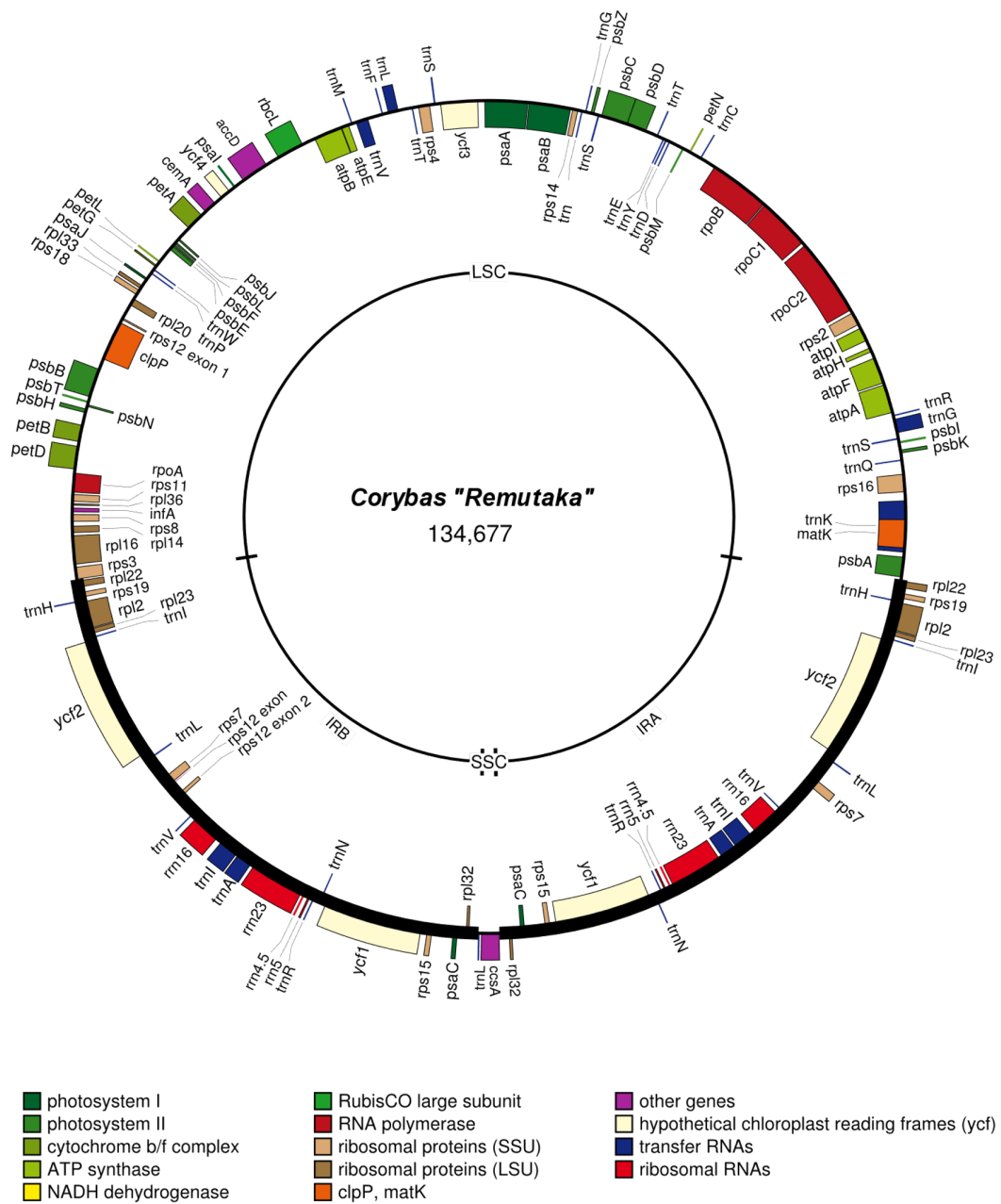
Supplementary Figure 5 The plastome of *Corybas dienemus* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).
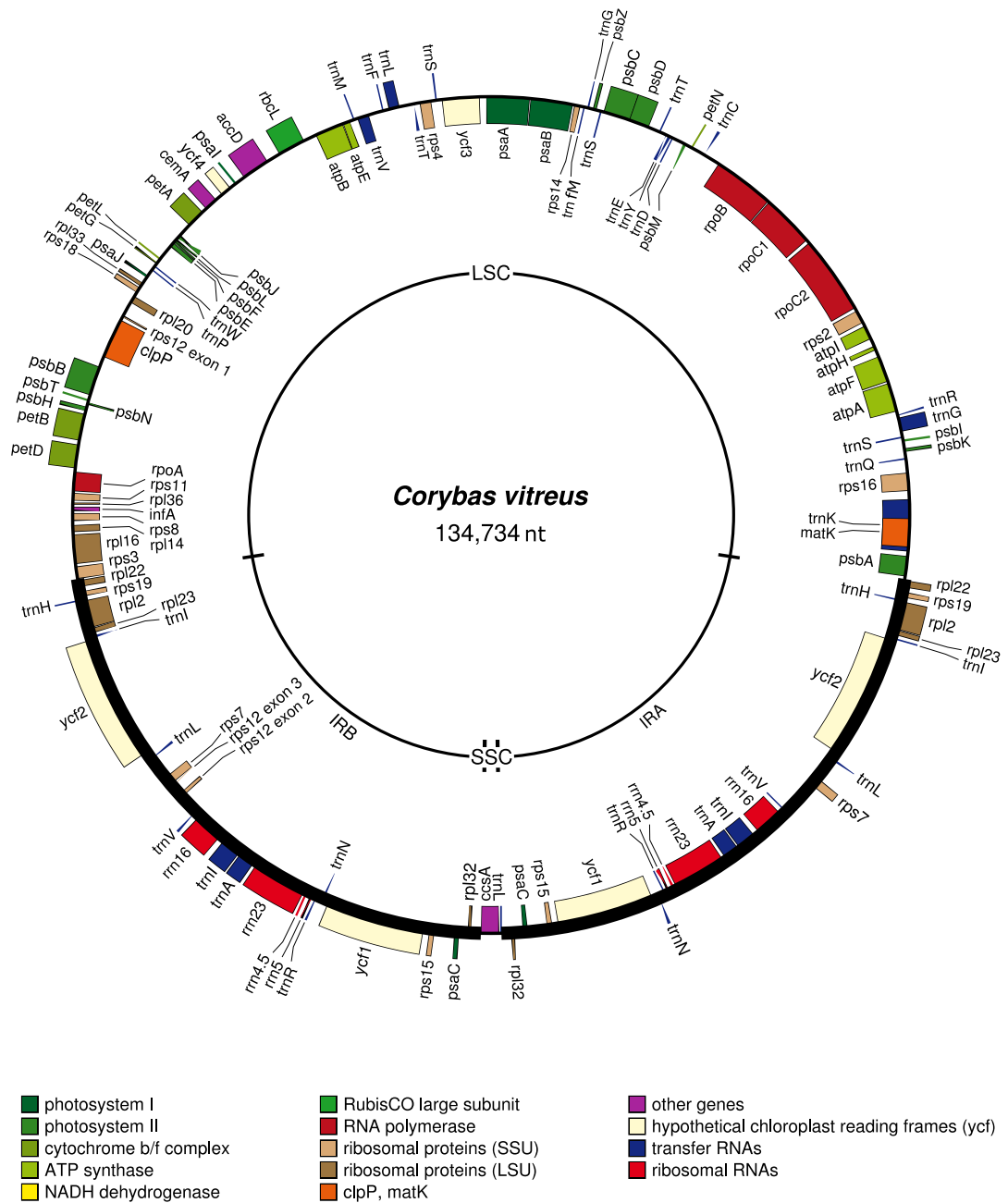
Supplementary Figure 6 The plastome of *Corybas iridescens* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).
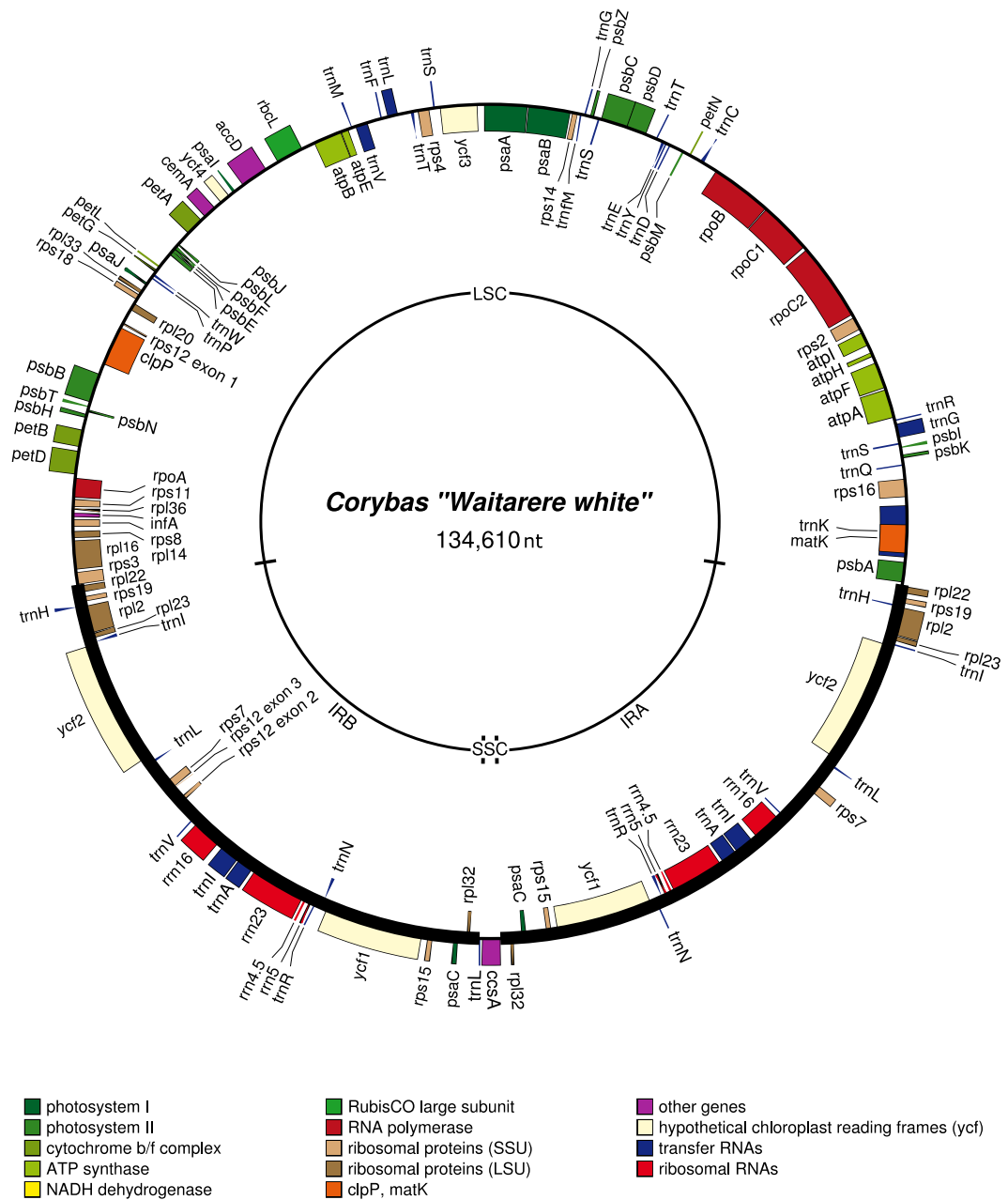
Supplementary Figure 7 The plastome of *Corybas macranthus* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).

Supplementary Figure 8 The plastome of *Corybas* "Rimutaka" Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR_A and IR_B), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).
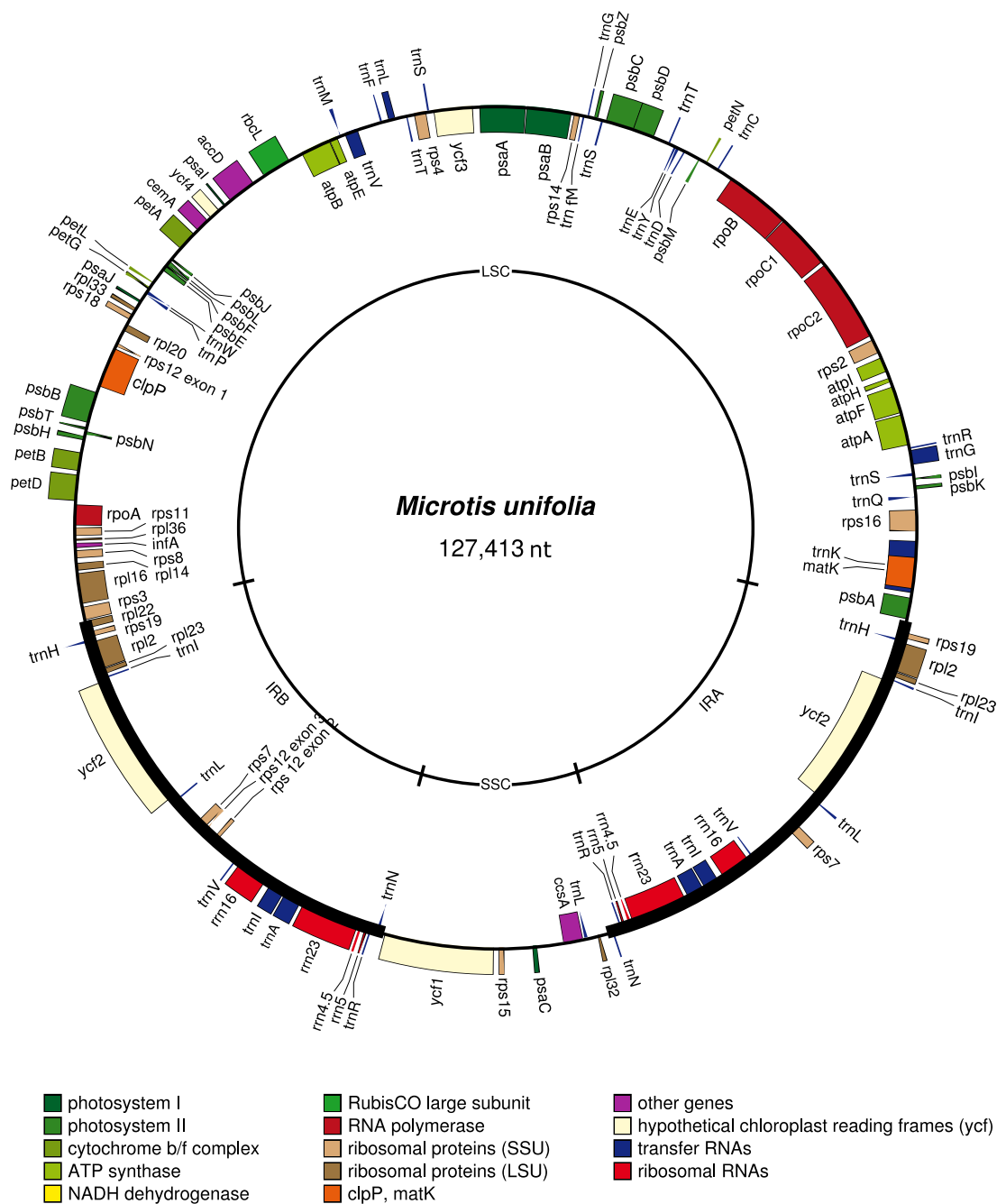
Supplementary Figure 9 The plastome of *Corybas vitreus* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013).

Supplementary Figure 10 The plastome of *Corybas* "Waitarere white" Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013)

Supplementary Figure 11 The plastome of *Microtis unifolia* Genes on the interior of the outer circle are transcribed in the forward direction; those to the exterior of this circle are transcribed in the reverse direction. The inner circle indicates extent of the inverted repeat (IR$_A$ and IR$_B$), the large single copy (LSC) region and the small single copy (SSC) region. The bold section on the outer circle indicates the IR regions. Plastome drawn using OrganellarGenomeDraw (Lohse et al., 2013)

Appendix III

Supplementary Table 2 Details of available heterotrophic plastomes between 54,000 and 84,000 nt in size

Species and authority	Family	Trophy	Plastome size (nt)	Number of putatively functional genes	Plastome structure	Genbank accession number	Reference
Orobanche latisquama F. W. Schultz	Orobanchaceae	Holoparasitic	80,361	52	Quadripartite, IR degraded	NC_02564	Wicke et al. (2013)
Epifagus virginiana (L.) Barton	Orobanchaceae	Holoparasitic	70,028	44	Quadripartite, IR intact	NC_001568	Wolfe et al. (1992)
Corybas cryptanthus Hatch (Eastbourne accession)	Orchidaceae	Mycoheterotrophic	69,376	52	One IR copy lost	MK883747	This thesis
Corybas cryptanthus (Taranaki accession)	Orchidaceae	Mycoheterotrophic	69,300	52	One IR copy lost	MK883746	This thesis
Orobanche gracilis Sm.	Orobanchaceae	Holoparasitic	65,533	56	Quadripartite, IR intact	NC_023464	Wicke et al. (2013)
Orobanche purpurea (Jacq.) Sojak	Orobanchaceae	Holoparasitic	62,891	52	Quadripartite, IR intact	NC_023132	Wicke et al. (2013)
Orobanche ramosa (L.) Pomel.	Orobanchaceae	Holoparasitic	62,304	54	IR absent	NC_023465	Wicke et al. (2013)
Orobanche aegyptiaca (Pers.) Pomel	Orobanchaceae	Holoparasitic	60,947	52	IR absent	KU212370	Wicke et al. (2016)
Orobanche lavandulacea (Rchb). Pomel	Orobanchaceae	Holoparasitic	60,878	51	IR absent	KU212371	Wicke et al. (2016)
Rhizanthella gardneri R.S. Rodgers	Orchidaceae	Mycoheterotrophic	59,190	36	Quadripartite, IR intact	NC_014874	Delannoy et al. (2011)

References

Delannoy, E., Fujii, S., Colas Des Francs-Small, C., Brundrett, M., & Small, I. (2011). Rampant Gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution, 28*, 2077-2086.

Wicke, S., Muller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., Renner, S. S., & Schneeweiss, G. M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell, 25*, 3711-3725.

Wicke, S., Müller, K. F., DePamphilis, C. W., Quandt, D., Bellot, S., & Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 9045-9050.

Wolfe, K. H., Morden, C. W., & Palmer, J. D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 10648-10652.