



A chromosome-scale *Gastrodia elata* genome and large-scale comparative genomic analysis indicate convergent evolution by gene loss in mycoheterotrophic and parasitic plants

Yuxing Xu^{1,†} , Yunting Lei^{1,†} , Zhongxiang Su¹, Man Zhao¹, Jingxiong Zhang¹, Guojing Shen¹, Lei Wang¹, Jing Li¹, Jinfeng Qi¹ and Jianqiang Wu^{1,2,*} 

¹Yunnan Key Laboratory for Wild Plant Resources, Department of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China, and

²CAS Center for Excellence in Biotic Interactions, University of Chinese Academy of Sciences, Beijing 100049, China

Received 10 August 2021; revised 27 September 2021; accepted 29 September 2021; published online 26 October 2021.

*For correspondence (email wujianqiang@mail.kib.ac.cn).

[†]These authors contributed equally to this work.

SUMMARY

Mycoheterotrophic and parasitic plants are heterotrophic and parasitize on fungi and plants, respectively, to obtain nutrients. Large-scale comparative genomics analysis has not been conducted in mycoheterotrophic or parasitic plants or between these two groups of parasites. We assembled a chromosome-level genome of the fully mycoheterotrophic plant *Gastrodia elata* (Orchidaceae) and performed comparative genomic analyses on the genomes of *G. elata* and four orchids (initial mycoheterotrophs), three parasitic plants (*Cuscuta australis*, *Striga asiatica*, and *Sapria himalayana*), and 36 autotrophs from various angiosperm lineages. It was found that while in the hemiparasite *S. asiatica* and initial mycoheterotrophic orchids, approximately 4–5% of the conserved orthogroups were lost, the fully heterotrophic *G. elata* and *C. australis* both lost approximately 10% of the conserved orthogroups, indicating that increased heterotrophy is positively associated with gene loss. Importantly, many genes that are essential for autotrophs, including those involved in photosynthesis, the circadian clock, flowering time regulation, immunity, nutrient uptake, and root and leaf development, were convergently lost in both *G. elata* and *C. australis*. The high-quality genome of *G. elata* will facilitate future studies on the physiology, ecology, and evolution of mycoheterotrophic plants, and our findings highlight the critical role of gene loss in the evolution of plants with heterotrophic lifestyles.

Keywords: *Gastrodia elata*, mycoheterotrophic plant, parasitic plant, gene loss, convergent evolution.

INTRODUCTION

In nature, most plants are autotrophs. However, mycoheterotrophic and parasitic plants partly or completely rely on other organisms to obtain nutrients, and many of them have low or even no photosynthetic capacity; furthermore, some of these heterotrophic plants have largely altered body plans and unique physiology, ecology, and evolutionary histories.

Due to very limited light availability, some plant species living under forest trees evolved to cheat their associated mycorrhizal fungi or neighboring free-living fungi and became parasitic on fungi; these unique plants are called mycoheterotrophic (Leake, 1994, 2004). Mycoheterotrophic plants evolved independently over 40 times in liverworts, monocots, and eudicots (Merckx et al., 2012). They may rely on fungi at their germination stage (initial

mycoheterotrophy); some are mixotrophs, which are photosynthetic and mycoheterotrophic at maturity (partial mycoheterotrophy); the fully mycoheterotrophic plants are achlorophyllous and are solely dependent on fungi during their entire life cycles (Leake, 1994).

Orchidaceae (Asparagales) is the one of the largest flowering plant family, comprising 750 genera and approximately 27 000 species (Chase et al., 2015; Li et al., 2019). More than 99% of Orchidaceae plants are at least initial mycoheterotrophs: Seeds of Orchidaceae plants are tiny (dust-like) and do not have endosperm, and certain fungi are required to provide the Orchidaceae seeds with nutrients for germination (McCormick et al., 2018). In addition to its great diversity in ecology and evolution, Orchidaceae is probably the most suitable plant family to study mycoheterotrophy. In Orchidaceae, more than 200 species from

at least 25 lineages are full mycoheterotrophs (Merckx et al., 2012), among which *Gastrodia elata* has been intensively studied for its medicinal effect and phytochemistry, given that it is a herbal medicine in eastern Asia (Zhan et al., 2016). Like most Orchidaceae species, *G. elata* seeds cannot germinate without obtaining nutrients from fungi, such as *Mycena dendrobii* or *Mycena osmundicola*. The symbiosis with *Mycena* fungi enables *G. elata* seeds to develop into a stage named protocorm, and gradually protocorms initiate parasitization on the saprophytic Armillaria fungi, and by exploiting the *Armillaria* fungi, *G. elata* protocorms gradually grow into vegetative propagation corms, immature tubers, and mature tubers; finally, a scape emerges from each mature tuber, and flowers are developed from the scape to produce seeds (Xu et al., 1989; Zhou, 1981). Throughout its life cycle, *G. elata* does not have a root or obvious leaves. The *G. elata*-associated *Mycena* spp. and *Armillaria* spp. have been isolated and cultured, allowing *G. elata* to be cultivated under agriculture and laboratory settings (Shimaoka et al., 2017). Although no genetic manipulation of either *G. elata* or its associated fungi has been reported, this mycoheterotrophic system is promising to become a model for studying the interaction and coevolution of mycoheterotrophic plants and the associated fungi.

About 1% of angiosperms from 12 or 13 independent lineages are haustorial parasitic plants (parasitic plants hereafter), which directly feed on other living plants using a special organ named haustorium (Westwood et al., 2010). Some parasitic plants, such as *Striga*, *Phelipanche*, *Orobanchaceae*, *Viscum*, and *Cuscuta*, are economically important, as they threaten agriculture or forestry (Clarke et al., 2019). The genus *Cuscuta* (dodder) is the only parasitic lineage in Convolvulaceae. Around 200 species of *Cuscuta* are globally distributed parasites. Given that *Cuscuta* species have no or very little photosynthetic activity, they are often considered to be holoparasites. Unlike many root parasites, germination of Cuscuta does not require chemical cues. *Cuscuta* seedlings rotate to search for potential hosts, and after contact with suitable adjacent objects, seedlings start to twin around the objects and haustoria are induced, and if the objects are compatible host plants, these initial haustoria penetrate the hosts to transfer water and nutrients; thereafter, *Cuscuta* starts to grow vigorously and many new stems are formed which further twin around hosts and new haustoria are developed along the contact regions between *Cuscuta* and host stems. Similar to *G. elata*, *Cuscuta* species do not have roots or obvious leaves.

The fully mycoheterotrophic and holoparasitic plants represent the advanced stages of parasitism evolution. Although evolved independently, mycoheterotrophic plants and parasitic plants have various physiological and morphological features in common. Both groups of plants experienced regressive evolution, given their largely

simplified body plans. Recently, several genomes of parasitic plants, including the holoparasite Cuscuta australis and *Cuscuta campestris* (Sun et al., 2018; Vogel et al., 2018), hemiparasite Striga asiatica (Yoshida et al., 2019), and the endoparasite Sapria himalayana (Cai et al., 2021), have been sequenced and analyzed. The *G. elata* genomes based on Illumina sequencing have also been released (Chen et al., 2020; Yuan et al., 2018). Here we assembled a chromosome-scale *G. elata* genome. Importantly, the genomes of *C. australis* and *G. elata* were compared with 42 genomes of plants, ranging from *Amborella* to monocots and eudicots, including the parasitic plants *S. asiatica* and *S. himalayana*. Our analyses indicate that gene loss is positively associated with the degree of heterotrophy, and we provide strong evidence of convergent evolution by gene loss in fully mycoheterotrophic and parasitic plants. We propose that in plants with a parasitic lifestyle, gene loss and its consequent physiological changes are likely adaptive and is thus positively selected for.

RESULTS

Sequencing, assembly, and annotation of the *Gastrodia elata* genome

Given that the previously built *G. elata* genome sequences have relatively short contigs, we used the PacBio Sequel II platform to sequence the *G. elata* genome, and Hi-C technology was employed for further scaffolding. *k*-mer analysis indicated that the *G. elata* genome is 1.092 Gb (Figure S1). The obtained long reads were assembled to contigs, whose total length is 1.043 Gb, covering 95% of the *G. elata* genome, and the contig N50 reached 21.33 Mb. The SNP and indel errors in the contigs were sequentially polished by the PacBio reads and corrected by the next-generation sequencing (NGS) reads (Table S1). After removal of contamination and organelle sequences (Figure S2), using the Hi-C chromatin contact data (Figure S3), we finally assembled the 18 *G. elata* chromosomes, on which 99.09% of the contigs were anchored (Figure 1). The total length of the non-anchored contigs is 9.45 Mb. By mapping the NGS reads to the genome assembly, we found that the mapping rate, error rate, and heterozygosity values are 97.89, 0.00094, and 0.00388, respectively (Table S1). Publicly available RNA sequencing (RNA-seq) data were also mapped to the genome assembly, obtaining mapping rates from 90.0 to 94.8% (Table S2).

Next the *G. elata* genome was annotated. Initially, based on structural information and homology, repetitive elements were identified, including telomeres, RNA genes, miniature inverted-repeat transposable elements, class I and II transposons, and various tandem repeats (Figure 1; Table S3). These repetitive elements occupied 66.36% of the *G. elata* genome. Annotation was performed using the

Hi-C 跟 scaffolding 有什麼關係?

要去查別人的

這個是什麼?

一個看

transposon types?

how many types of parasitic plants

葛縷子屬

旋花科

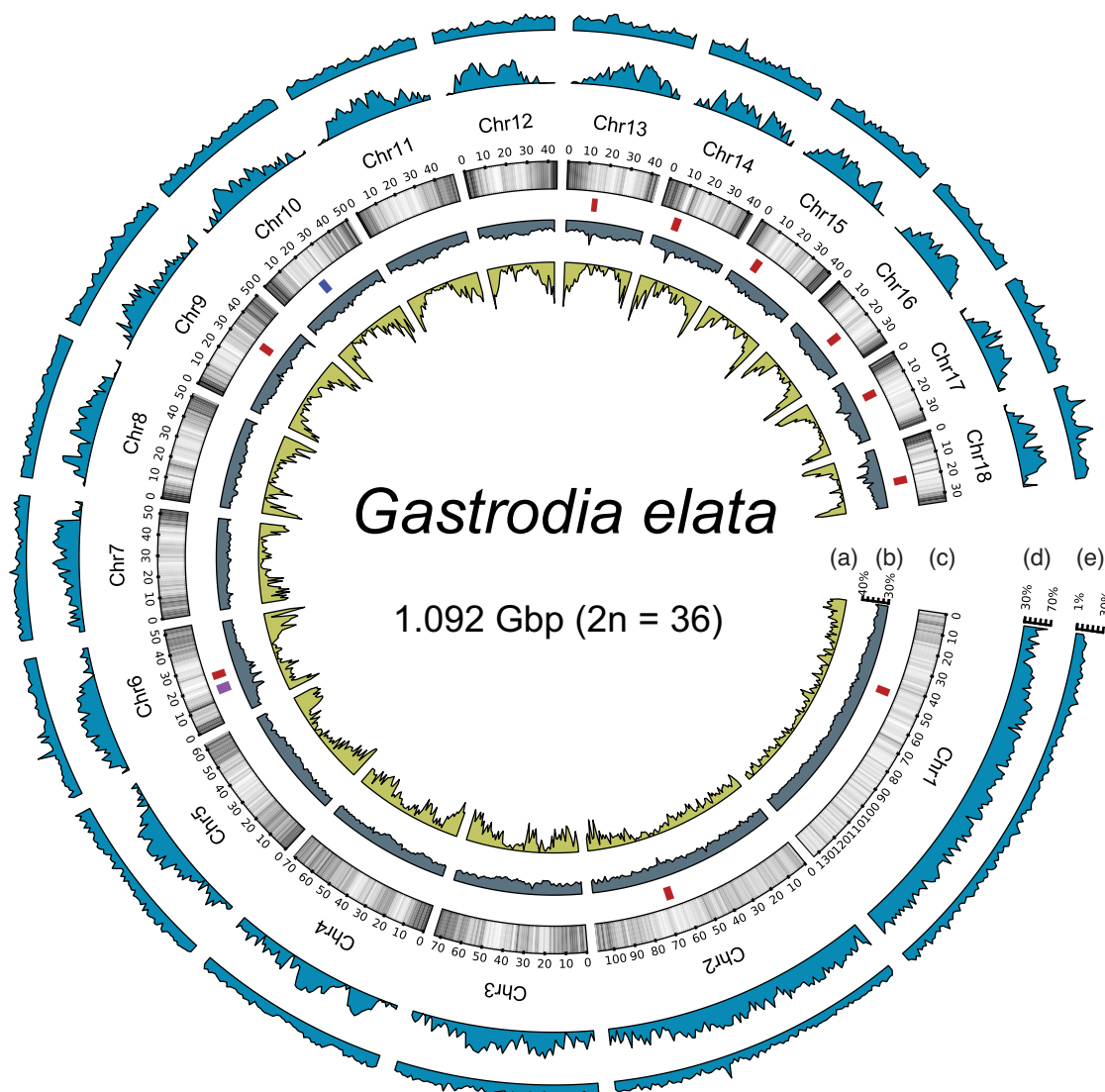


Figure 1. Overview of the *Gastrodia elata* genome assembly.

(a) Genome-wide distribution of relative expression levels of all *G. elata* transcripts. (b–e) GC content (b), gene density (c), and densities of Class I retrotransposons (d) and Class II DNA transposons (e). The red, purple, and blue rectangles indicate the positions of centromere-specific satellites, 45S rDNA, and 5S rDNA, respectively.

植物是45 s 哦

published RNA-seq data (from *G. elata* protocorms, juvenile tubers, immature and mature tubers, and scapes), homologous genes from the closely related species, and results from the *de novo* prediction. The annotations from Yuan et al. (2018) were further merged with those from this work, generating 21 115 protein-coding genes and 3664 pseudogenes (Table S3). The *G. elata* genome from Yuan et al. (2018) has 86.60% coding genes identical to the updated genome and 11.50% were structurally modified; only 361 genes (1.93%) from Yuan et al. (2018) were not found in the new *G. elata* genome assembly and 2941 genes were newly identified (Figure S4). Benchmarking universal single-copy orthologs (BUSCO) analysis

indicated that the completeness of annotation is 76.40% (Table S4).

Gene families in *G. elata*

Although the ancestor of Orchidaceous plants experienced a whole-genome duplication event (Zhang et al., 2017), the genomes of all orchids that have been sequenced thus far do not contain many protein-coding genes. Except the *Vanilla planifolia* genome, which has 29 044 genes, all reported genomes of orchids *Dendrobium catenatum* (Zhang et al., 2016), *Phalaenopsis equestris* (Cai et al., 2015), and *Apostasia shenzhenica* (Zhang et al., 2017) possess less than 22 000 genes, while

autotrophic angiosperms normally have more than 25 000 genes (Table S5).

Next, gene families in *Orchidaceae* and *Asparagus officinalis* (basal species of Asparagales) and representative species from all other orders in monocots were selected for gene family expansion and contraction analysis, and species from *Magnoliids* and *Amborella trichopoda* (basal species from angiosperms) were used as the outgroups (Table S6). We found that the orchid ancestor experienced contraction in 616 gene families, and except *V. planifolia*, approximately 75% of the conserved gene families in all orchids have sizes below the average gene family size of all species examined (Figure 2; Table S6). Notably, *G. elata* exhibited the largest contraction: 1440 *G. elata*-specific gene families were contracted (Figure 2; Table S6).

Introns in *G. elata*

In the endoparasitic plant *S. himalayana*, introns are generally longer than in autotrophs, reaching an average of 1527 bp, and the long introns in *S. himalayana* were hypothesized to be caused by relaxed selection, due to its parasitic lifestyle (Cai et al., 2021). The genomes of orchids *P. equestris* and *A. shenzhenica* were also found to have large introns (Cai et al., 2015; Zhang et al., 2017). Therefore, to explore whether there is a correlation between

mycoheterotrophy and increased intron size, we also inspected the genome of *G. elata*.

Even though the *G. elata* genome harbors a small number of genes, the total length of genes reaches 303.5 Mb, exceeding the whole-genome size of many plants, 84.3% of which constitutes intron regions (Table S7). The introns in the *G. elata* genome are much larger than those in most autotrophic plants: 21% of its introns are longer than 5 kb, whereas in rice (*Oryza sativa*), only 0.2% of the introns are longer than 5 kb, and even in maize (*Zea mays*), which has a genome larger than that of *G. elata*, this value reaches only 1.89% (Table S7). Introns longer than 5 kb can be found in 38.6% of *G. elata* genes (Table S7). Similarly, long introns were also found in other *Orchidaceae* plants, except the basal orchid *A. shenzhenica*; notably, the *G. elata* genome showed the strongest pattern of intron elongation (Figure 3a). We next examined whether intron numbers increased in orchids, but no big differences were found between the densities of introns in *Orchidaceae* and other plants (Figure 3b). If introns longer than 5 kb are considered to be large introns, in *V. planifolia* and *Epidendroideae* plants, the total lengths of large introns are more than 70% of the total intron lengths, much higher than in the closely related species *A.*

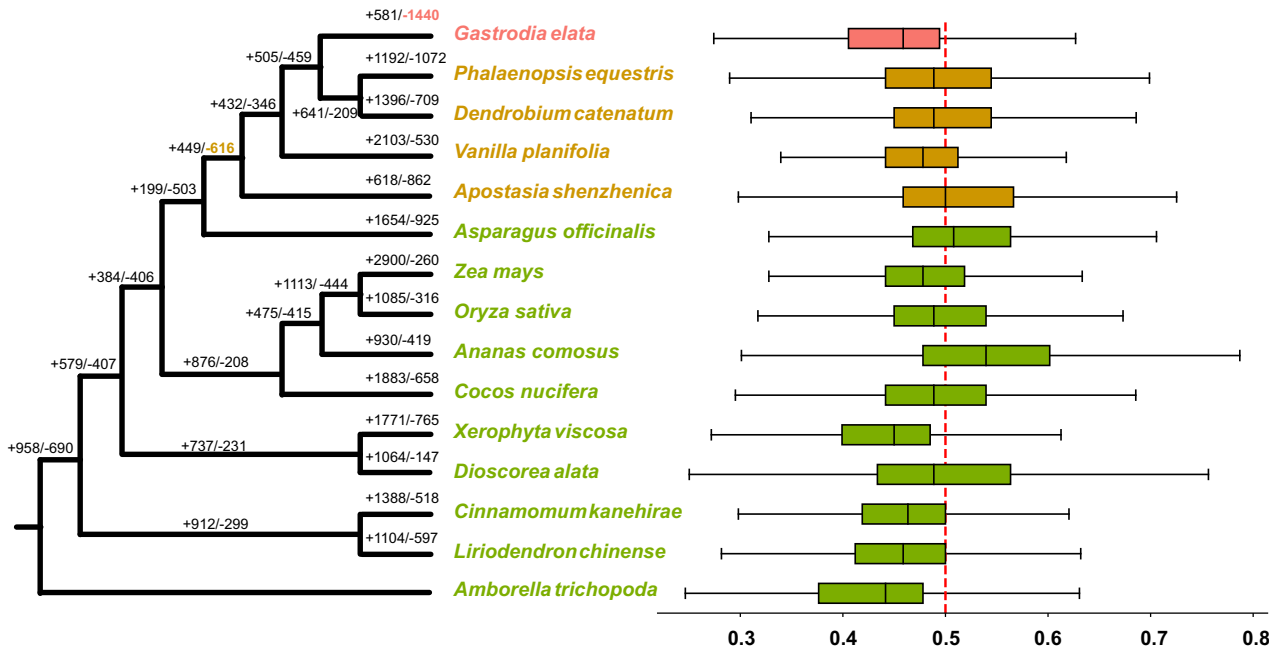


Figure 2. Gene family expansion and contraction in *Orchidaceae* and outgroups. Left panel: Significantly expanded and contracted gene families. The numbers above the branches indicate the expanded (before slash) and contracted (after slash) gene families. Right panel: Tukey boxplot overview of the differences among the gene numbers of the conserved gene families, based on the *F*-index values (*F*-indices range from 0 to 1; when $F = 0.5$, $F < 0.5$, or $F > 0.5$, the gene number in the given gene family is equal to, smaller than, or greater than the average size of this gene family in all species). The left and right sides of the boxes are the first and third quartiles, respectively; medians of the data are shown as the bands in the boxes; for each box, the whiskers represent the smallest and biggest data that are still within 1.5 times the interquartile range of the lower and upper quartiles.

shenzhenica and *A. officinalis* (approximately 50%) (Figure 3c). Our analysis indicated that at least 30% of these large introns in all orchids analyzed are transposon sequences (Figure 3c; Table S7). To investigate how the large introns expanded specifically in Orchidaceae, we compared the single-copy orthologs in orchids, including all the sequenced orchids and closely related monocots, to identify 'orthologous introns' whose positions are conserved in the corresponding genes. Indeed, these orthologous introns are generally larger in orchids than in other monocots (Figure 3d). Assuming non-orchid monocot plants retained the ancestral intron sizes, we inferred that introns smaller than 200 bp mostly did not expand, whereas those bigger than 200 bp tended to lengthen (Figure 3d). We speculate that the expansion of introns in orchids was due to accumulation of different types of transposons over time; most introns smaller than 200 bp remained, as they are too small to have enough chance of being inserted by transposons.

orthologous introns
怎麼抓出來的
又是神奇演算法嗎
(方法應該會講?)

Gene loss in mycoheterotrophic and parasitic plants

Large-scale gene loss has been detected in various parasitic plants (Cai et al., 2021; Sun et al., 2018; Vogel et al., 2018; Yoshida et al., 2019). Furthermore, even though contraction at the gene family level was detected in the genomes of *G. elata* and other orchid species, a contracted gene family does not necessarily have impaired biological functions that are normally carried out by the genes of this family, unless the gene family lost its all members. Therefore, we next specifically compared the lost genes in both mycoheterotrophic and parasitic plants with varying degree of heterotrophy.

Orchidaceous mycoheterotrophic species *A. shenzhenica* (Zhang et al., 2017), *V. planifolia* (Hasing et al., 2020), and *G. elata* and the parasitic plants *S. asiatica* (Yoshida et al., 2019), *S. himalayana* (Cai et al., 2021), and *C. australis* (Sun et al., 2018) were compared with 36 autotrophic plants from all lineages of angiosperms (Table S5). Among these

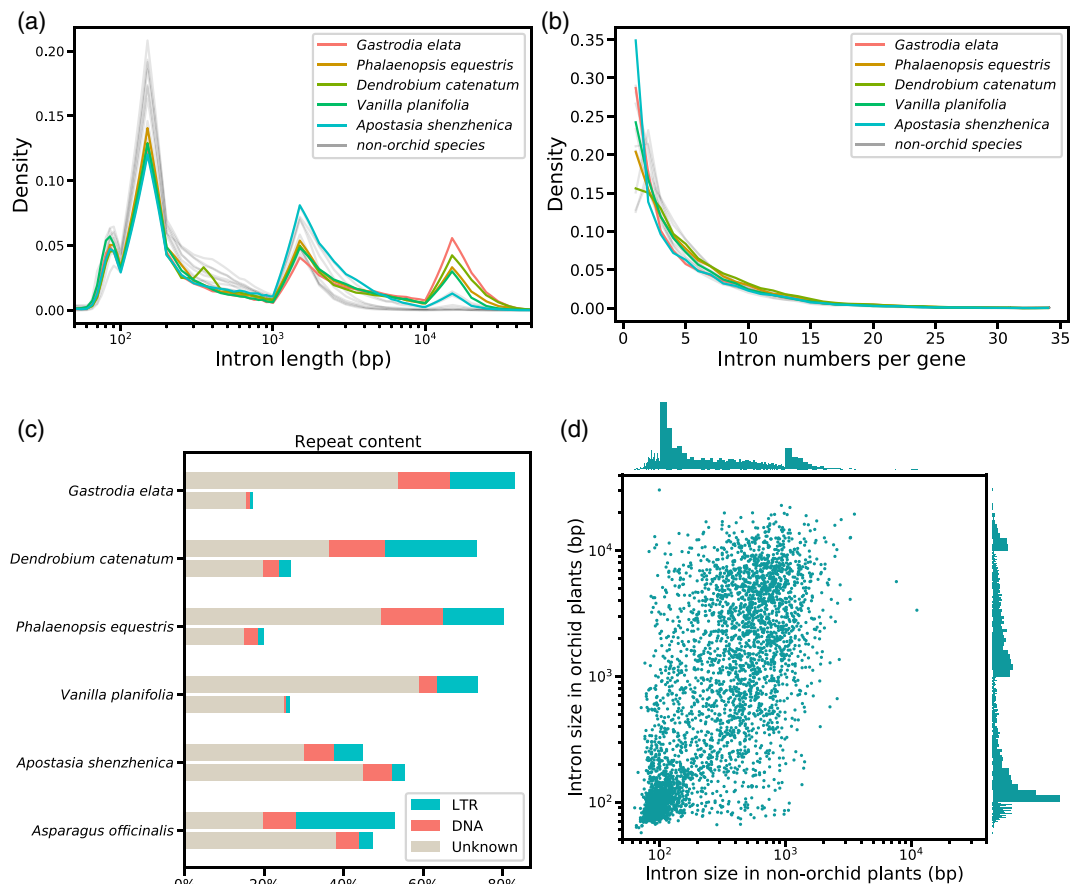


Figure 3. Intron properties of *Gastrodia elata* and other Orchidaceae species.

(a,b) Distribution of intron sizes (a) and numbers of introns (b) in all genes. Colored lines indicate Orchidaceae plants and gray lines indicate non-orchid plants (species are listed in Table S7). (c) Contents of repetitive elements in the introns of Orchid plants and *Asparagus officinalis*. For each species, the top bar represents introns bigger than 5 kb and the bottom bar represents introns equal to or smaller than 5 kb. Strong cyan indicates long terminal repeat (LTR) retrotransposons, light red indicates DNA transposons, and grayish orange represents unknown sequences. (d) Size distribution of introns that are conserved in orchids and non-orchid species.

phylogenetically distant species, 10 236 orthogroups (OGs) were found to be well conserved. Overall, the initial mycoheterotrophic orchids exhibited a similar scale (approximately 4.5%) of gene loss as the hemiparasitic plants (*A. shenzhenica*: 463/4.52%; *V. planifolia*: 528/5.16%; *S. asiatica*: 402/3.93%; first numbers indicate the numbers of lost OGs, the percentages after the slash are the ratios between lost and conserved OGs) (Table S8; Figure S5). When plants evolved to be completely heterotrophic, much greater levels of gene loss were detected: Both *G. elata* and *C. australis* showed approximately 10% gene loss (*G. elata*: 1211/11.83%; *C. australis*: 1012/9.89%) (Table S8). Strikingly, consistent with its extremely simplified body plan, the endoparasite *S. himalayana* lost the largest number of genes (3604/35.21%) (Table S8; Figure S5).

Next, in order to gain insight into the convergent evolution in these heterotrophic plants, we compared the lost genes between mycoheterotrophic and parasitic plants. Comparison between the initial mycoheterotroph *A. shenzhenica* and the hemiparasite *S. asiatica* indicated that only 74 OGs (comprising 15.98 and 17.62% of the lost genes in *A. shenzhenica* and *S. asiatica*, respectively) were in common (Table S9). By contrast, 35.60 and 42.59% of the lost genes in the full mycoheterotroph *G. elata* and the holoparasite *C. australis* (461 OGs) were in common, indicating convergent evolution by gene loss in these two fully heterotrophic plants. Remarkably, approximately 65–80% of the lost OGs in all other heterotrophic species were found to be lost in *S. himalayana* (Table S9). It seems that most of the gene loss events in partial heterotrophic plants are lineage-specific, while in fully heterotrophic plants, many more gene loss events were driven by convergent evolution. Our comparison also indicated that *S. himalayana* is at the extreme end of gene loss in heterotrophic plants, which recapitulated most of the gene loss events that took place in mycoheterotrophic and parasitic plants.

To further investigate the convergent evolution by gene loss in both mycoheterotrophic and parasitic plants, we performed function enrichment of the commonly lost genes between *G. elata* and *C. australis*; *S. himalayana* was excluded from further analysis, since the extent of its gene loss is too extreme. It was found that the functional categories photosynthesis, nutrient uptake, and external stimuli response were the most significant (Table S10). Next, MapMan gene functional categories photosynthesis, organelle machinery, nutrient uptake, and external stimuli response were retrieved to obtain the genes in each category, and these genes were further filtered using the criterion that a valid gene should exist in both autotrophic plants *A. officinalis* and *Arabidopsis* (Figure 4; Table S11). These gene sets were designated as the backgrounds for analyzing gene loss in two groups of plants, group 1 (*A. officinalis*, *A. shenzhenica*, and *G. elata*, for mycoheterotrophic

plants and their closely related species) and group 2 (*Arabidopsis*, *S. asiatica*, and *C. australis*, for parasitic plants and their closely related species). The numbers of lost genes in each species were divided by the numbers of total genes in each category to reveal the degrees of gene loss in these functional categories. We found an increasing trend of gene loss from *A. officinalis* to *A. shenzhenica* and *G. elata*; especially *G. elata* displayed the highest level of gene loss in all categories (Figure 4). Compared with the hemiparasite *S. asiatica* and the autotroph *Arabidopsis*, *C. australis* also exhibited the highest levels of gene loss in all categories, even though in the categories photosynthesis and nutrient uptake, the numbers of lost genes in *C. australis* were not very different from those in *S. asiatica* (Figure 4).

Moreover, in some categories, mycoheterotrophic plants and parasitic plants have different patterns of gene loss, e.g., in the categories carbohydrate metabolism and coenzyme metabolism, initial and full mycoheterotrophic orchids lost more genes than did the hemiparasite *S. asiatica* and the holoparasite *C. australis*; by contrast, in the symbiosis signaling pathway, parasitic plants exhibited large-scale losses, but no gene loss happened in mycoheterotrophic plants (Figure 4).

Pathways affected by gene loss in *G. elata* and *C. australis*

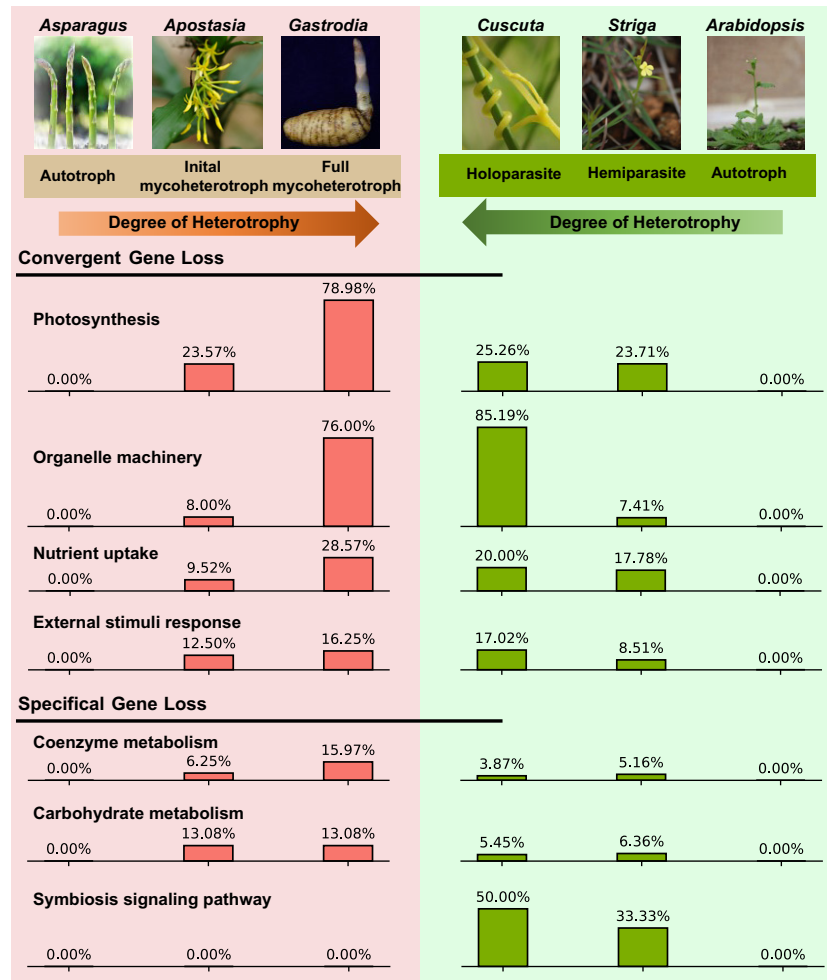
Next, to better understand the signaling and metabolic pathways affected by gene loss, we specifically inspected the genes lost in *G. elata* and *C. australis*. The genomes of 14 species, including *G. elata*, *C. australis*, and *Arabidopsis*, were used for phylogenomic analysis, so that the orthologs of all *Arabidopsis* genes in *G. elata* and *C. australis* could be obtained (Table S9).

Plastome- and photosynthesis-related genes. Consistent with its achlorophyllous morphology, *G. elata* does not have any photosynthetic activity, as indicated by analysis of the maximum efficiency of photosystem II photochemistry (Fv/Fm) (Figure S6), and *G. elata* has an extremely reduced number of plastome genes (73%) (Table S12). *Cuscuta australis* still retains very little but detectable photosynthetic activity (Sun et al., 2018), and compared with *G. elata*, fewer genes were lost in the *C. australis* plastome (22%) (Table S12). Almost none of the conserved photosynthesis-related genes were found in *G. elata*: Among 64 conserved photosynthesis-related genes, only *pasA* (photosystem I) and *FD3*, *FD4*, *RFNR1*, and *FRNR2* (photosynthetic electron transport) were found in *G. elata* (Table S13). By contrast, *C. australis* retained many more photosynthesis-related genes (45 among 64 conserved photosynthesis-related genes) (Table S13).

Light perception, circadian clock, and flowering time regulation. Phytochromes *phyB*, *phyD*, and *phyE*, cryptochromes

Figure 4. Overview of gene loss patterns in myco-heterotrophic and parasitic plants. *Asparagus*, *Apostasia*, and *Gastrodia*, which are an autotroph, an initial mycoheterotroph, and a full myco-heterotroph, respectively, and *Arabidopsis*, *Striga*, and *Cuscuta*, which are an autotroph, a hemiparasite, and a holoparasite, respectively, were analyzed for their gene loss. The upper and lower panels show the convergent and specific pattern of gene loss in mycoheterotrophic and parasitic plants, respectively. The percentage labeled on top of each bar indicates the ratio of the lost genes to all the conserved genes of the given functional category in autotrophic plants.

MapMan分類的嗎



CRY1 and *CRY2*, and the *ZTL* (*ZEITLUPE*) family member *FKF1* (Sanchez et al., 2020) were lost in the *G. elata* genome (Figure 5a; Table S14). Similarly, *TOC1*, *PRR5*, and *PRR9*, which encode important clock components (Greenwood and Locke, 2020) are also absent in *G. elata* (Figure 5a; Table S14). Among these, only *FKF1*, but not the others, is absent in the *C. australis* genome (Figure 5a; Table S14). These results are consistent with the fact that *G. elata* spends more than 2 years underground as a tuber and thus light and circadian signals are unlikely needed for *G. elata* growth and development. In *G. elata*, in addition to loss of *TOC1*, *CCA1* did not show any transcriptional changes (Figure 5b), indicating a defective circadian system. By contrast, even though *Cuscuta* has very limited photosynthetic activity, light perception in *Cuscuta* plays an important role in sensing the existence of neighboring plants for parasitization (Smith et al., 2021), and consistently, quantitative PCR (qPCR) analysis confirmed the oscillation of *CCA1* expression (Figure 5b).

Flowering is a complex process that is controlled by multiple pathways (Blumel et al., 2015). Previously it was

found that the *C. australis* genome lost many important flowering regulatory genes, including those involved in the vernalization, autonomous, photoperiod, and circadian pathways (Sun et al., 2018). In *G. elata*, the vernalization and autonomous pathways are retained; however, the circadian clock and photoperiod pathways are very likely non-functional (Figure 5a; Table S14). In addition to the finding that the circadian genes *TOC1* and *PRR5/9* are missing, the *G. elata* genome also lost several photoperiod genes, including *CO*, which is a hub for signal integration of the circadian clock and photoperiodic flowering, and *CDF1* to *CDF3*, which encode transcriptional repressors of *CO* (Figure 5a; Table S14). Given that the aging pathway of *G. elata* remains intact (Figure 5a; Table S14), it is plausible that flowering could be controlled by the aging pathway in *G. elata*, which is related to sugar signaling (Yu et al., 2013).

Immunity. We specifically inspected genes involved in salicylic acid (SA) biosynthesis and signaling, given that SA is important for disease resistance (Ding and Ding,

在地下待兩年跟後面的敘述接不起來?

為什麼要用qPCR?

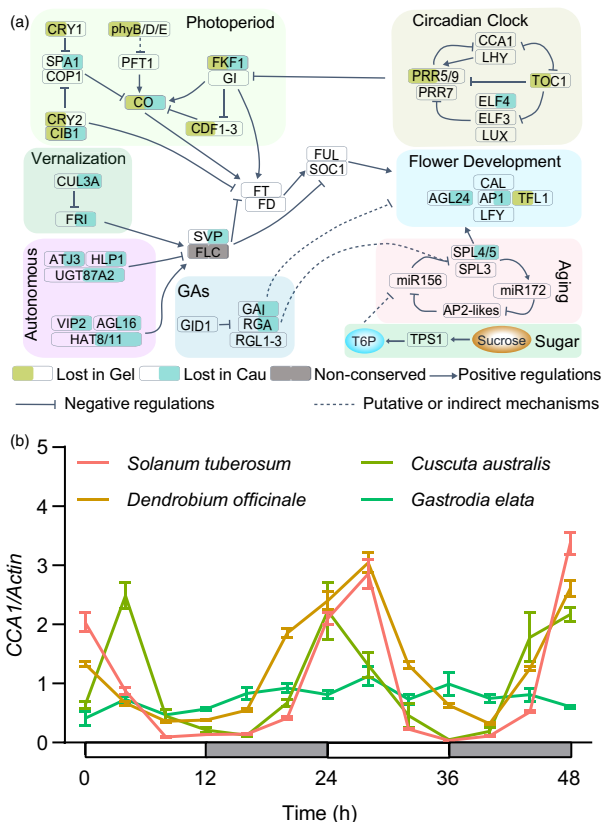


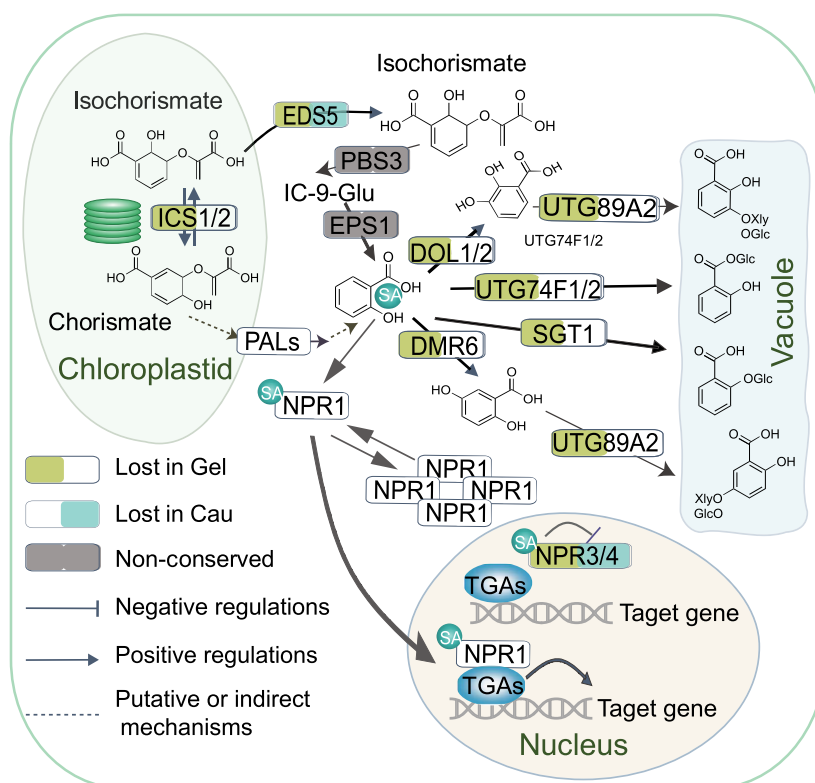
Figure 5. Lost genes involved in regulation of plant flowering time. (a) Simplified gene network controlling flowering time. Genes with a moderate yellow background on the left or a very soft cyan background on the right are lost in *G. elata* (Gel) or *C. australis* (Cau), respectively; genes with a gray background are not conserved. (b) The diurnal expression of CCA1 in *S. tuberosum*, *D. officinale*, *C. australis*, and *G. elata*. The white and gray time periods on the x-axis indicate day and night, respectively. The data are presented as the mean \pm SD ($n = 5$).

2020). *ICS1/2*, encoding isochorismate synthases, and *EDS5*, encoding a MATE family transporter, which are all involved in SA biosynthesis, are missing in the *G. elata* genome (Figure 6; Table S15). Consistently, when we infected *G. elata* tubers with *Pseudomonas syringae*, SA levels did not show any changes, while in the orchids *P. equestris* and *D. catenatum*, SA levels increased by 13 and 157%, respectively (Figure S7). *NPR3* and *NPR4*, encoding SA receptors, are missing in *G. elata* (Figure 6; Table S15). Notably, *EDS5*, *NPR3*, and *NPR4* are also absent in *C. australis* (Figure 6; Table S15). The important downstream genes of SA signaling, such as *EDS1*, *PAD4*, *ALD1*, and *FMO1*, are all missing in *G. elata* and *C. australis* (Figure 6; Table S15). It is not yet known how *G. elata* responds to pathogens, although it has been reported that anti-microbe peptides may play a role in the immunity of *G. elata*, such as *Gastrodia* antifungal protein (Hu et al., 1988; Wang et al., 2001). 有料X

Nutrient uptake and transport. Since *G. elata* does not have roots, we suspected that *G. elata* may also have lost genes which function in nutrient uptake, like *C. australis* (Sun et al., 2018). Indeed, the *G. elata* genome does not contain the high-affinity nitrogen transport genes *NRT2.1*, *2.2*, *2.4*, and *2.5* (Wang et al., 2018), which are normally expressed in roots (Table S16). Notably, these transporters are also missing in the *C. australis* genome (Sun et al., 2018). Moreover, *NPF2.7*, encoding a transporter functioning in nitrate release out of roots under acidic conditions (Wang et al., 2018), was lost in *C. australis* (Table S16). Similarly, *NPF4.6*, encoding a low-affinity nitrate transporter, and *NPF1.1* and *NPF1.2*, which function in redistributing xylem-borne nitrate into developing young leaves (Wang et al., 2018), were not found in *G. elata* (Table S16). Notably, *SKOR* and *GORK*, which mediate K^+ release from root parenchyma cells into the xylem for root-to-shoot K^+ transport and mainly control K^+ -selective outward-rectifying channels, respectively (Wang et al., 2021), are missing in *G. elata* (Table S16). Even though *SKOR* and *GORK* are retained in *C. australis*, several genes important for regulating K^+ uptake and transport were lost (Wang et al., 2021), including *CIPK6*, *HAK5*, *CBL1*, *CBL4*, *CBL9*, *NHX1*, and *NHX2* (Table S16). Phosphate signaling genes are also lost to certain degrees in *G. elata* and *C. australis* (Table S16): In *G. elata*, *SPX1* and *SPX2*, which regulate phosphate homeostasis (Wang et al., 2014), were found to be missing; in *C. australis*, *PHT1;1* to *PHT1;3*, which play important roles in phosphate uptake from the rhizosphere into plant roots (Wang et al., 2021), were lost.

Leaf and root development. Given the leaf- and rootless morphology of *G. elata*, we specifically inspected genes important for leaf and root development. *KNOX* genes were specifically inspected, as they regulate leaf initiation (Du et al., 2018). *KNAT5* and *KNAT7* are missing in *G. elata* (Table S17); in *C. australis*, *KNAT2*, 3, 4, and 6 were deleted (Table S17). Furthermore, *SPCH* and its closest paralog *MUTE*, which encode key transcriptional switches for the stomatal patterning ligand-receptor systems (Han et al., 2018), are retained in *C. australis*, whereas they were lost in *G. elata* (Table S17). In *C. australis*, *WOX5* and *PTL1*, 2, 4, 5, and 6, which are master regulators of root organogenesis (Motte et al., 2019), are missing (Sun et al., 2018); *WOX5* was found in *G. elata*, but *PTL1-3* and *5-7* were also lost (Table S17). *PIN2* and *AUX1*, auxin signaling genes which are essential for root gravitropism control (Motte et al., 2019), are absent in the *C. australis* and *G. elata* genomes, respectively (Table S17). Several genes functioning in lateral root development (Motte et al., 2019), *SLR*, *LBD16*, *ARF7*, *ARF19*, *AFB3*, and *LZY1*, are missing in *G. elata*, and similarly the *C. australis* genome does not contain *MAKR4*, *IAA28*, *LBD16*, *GATA23*, *AFB3*, *LZY2*, and *LZY4* (Table S17). many genes lost

Figure 6. Gene loss in the salicylic acid pathway. Simplified gene network of salicylic acid metabolism and signaling in *G. elata* (Gel) and *C. australis* (Cau). Genes with a moderate yellow background on the left or a very soft cyan background on the right are lost in *G. elata* or *C. australis*, respectively; genes with a gray background are not conserved.



Symbiosis signaling pathway. The receptor-like kinase SymRK, the calmodulin-dependent protein kinase CCaMK, and the transcription factor CYCLOPS are important components of the common symbiosis signaling pathway (Oldroyd, 2013), and the genes encoding these proteins are all missing in *C. australis*, while *CCaMK* and *CYCLOPS* are retained in *G. elata* (Figure 7; Table S18). The GRAS transcription factor RAD1 and the half-ATP-binding cassette (ABC) transporters STR and STR2 are essential for functional arbuscular mycorrhizal symbiosis (Jiang et al., 2017; Luginbuehl et al., 2017). We found that *RAD1*, *STR*, and *STR2* are all missing in *G. elata* and *C. australis* (Figure 7; Table S18). VAPYRIN, LIN, and SYN function in the formation of structures required for the intracellular accommodation and infection of arbuscular mycorrhizal symbiosis, root-nodule symbiosis, and mycorrhizal symbiosis (Huisman et al., 2016; Liu et al., 2019; Murray et al., 2011). In *G. elata*, *VAPYRIN*, *LIN*, and *SYN* are retained, while all these three genes are missing in *C. australis* (Figure 7; Table S18).

Vitamin biosynthesis. Phylloquinol is the precursor of phylloquinone (vitamin K1, functions as a coenzyme) (Basset et al., 2017), and we found that almost all genes encoding enzymes involved in phylloquinol biosynthesis are absent in the *G. elata* genome (Figure S8a; Table S19), suggesting a defect of vitamin K1 biosynthesis in *G. elata*.

The enzyme MenG, which converts demethylphyloquinol to phyloquinol, does not exist in either *C. australis* or *G. elata* (Figure S8a; Table S19). NCS1 is required for thiamine (vitamin B1, another coenzyme) and thiamine diphosphate biosynthesis and transport (Beaudoin et al., 2018; Noordally et al., 2020), and it is absent in both *G. elata* and *C. australis*; we also noticed that the genes encoding the thiamine biosynthesis enzymes THIC and THI1 (Kong et al., 2008; Machado et al., 1996) are also missing in *G. elata*, but retained in *C. australis* (Figure S8b; Table S19). Therefore, it was expected that *G. elata* and *C. australis* may not contain phyloquinone or thiamine. However, HPLC-MS/MS analysis detected both vitamins in *C. australis* and *G. elata* (Figure S8c,d). We hypothesize that phyloquinone or thiamine is transferred from *Armillaria*; supporting this possibility, both vitamins were detected in *Armillaria rhizomorph* (Figure S8c,d). Similarly, *C. australis* supposedly obtains these vitamins from the host plants.

Methylerythritol phosphate pathway. Strikingly, genes encoding DXSs, which catalyze the formation of 1-deoxy-D-xylulose 5-phosphate (DXP) (Estevez et al., 2001), are missing in *G. elata* (Figure S9a; Table S19). DXP is the upstream precursor for the production of geranylgeranyl pyrophosphate (GGPP), which is converted to various important metabolites, including terpenes, carotenoids, and even the hormone abscisic acid (ABA) and

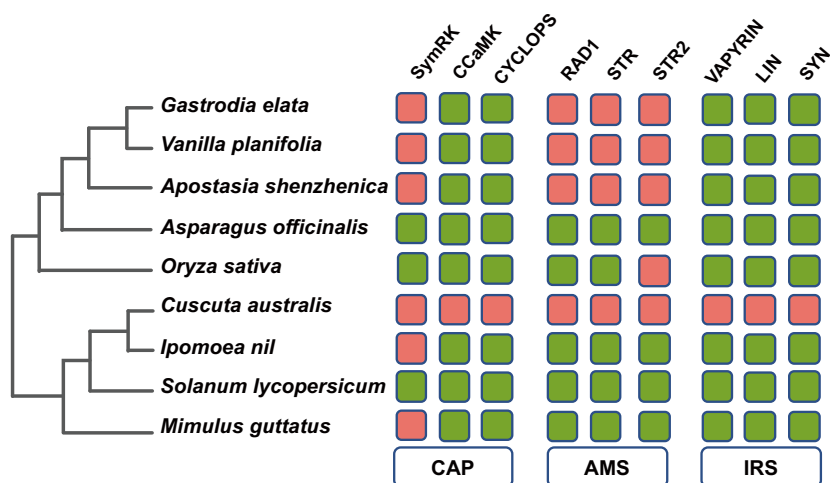


Figure 7. Gene loss in symbiotic genes.

The common symbiosis pathway (CSP) genes *SymRK*, *CCaMK*, and *CYCLOPS* are conserved in most land plants. *RAD1*, *STR*, and *STR2* are conserved exclusively in species with arbuscular mycorrhizal symbiosis (AMS). The infection-related symbiosis (IRS) genes *VAPYRIN*, *SYN*, and *LIN* are specific to angiosperm species forming intracellular symbiosis. The red and green square blocks indicate absence and presence, respectively, of the ortholog genes in the corresponding plant of the left species tree.

strigolactones (Nisar et al., 2015). We also found that genes encoding LUT2, LCY1, LUT5, and LUT1, which sequentially convert all-*trans*-lycopene to lutein, are all absent in *G. elata* (Figure S9a; Table S20). Given that β -carotene is the precursor of ABA and LCY1 is required for β -carotene biosynthesis (Nambara and Marion-Poll, 2005), it is plausible that *G. elata* cannot synthesize ABA (Figure S9a; Table S20). By contrast, these abovementioned genes in the methylerythritol phosphate (MEP) pathway are retained in *C. australis* (Figure S9a; Table S20). Using HPLC and HPLC-MS/MS, we quantified the contents of lutein/zeaxanthin, β -carotene, and ABA in *G. elata*. Compared with *Arabidopsis* and *D. catenatum* (positive controls), *C. australis* had much lower contents of lutein, zeaxanthin, and β -carotene, and these compounds were undetectable in *G. elata* (Figure S9b,c); however, ABA was detected in *G. elata*, although the content was only approximately 10% of that in the fungal host *Armillaria* (Figure S9d). It is possible that *G. elata* is able to transfer ABA directly from *Armillaria* for its own use.

DISCUSSION

Mycoheterotrophic and parasitic plants, especially the fully mycoheterotrophic and holoparasitic plants, constitute two groups of intriguing plants with parasitic lifestyles. Little is known about their physiology and ecology, including how they interact with the fungal or plant hosts and how their physiology enables them to adapt to their parasitic lifestyles. To understand these unique plant life forms, it is particularly important to know how parasitism in these plants evolved. There are 12 to 13 different lineages of parasitic plants (Westwood et al., 2010). Mycoheterotrophic plants have independently evolved more than 40 times (Merckx et al., 2012). Due to the limited availability of genomes of parasitic and mycoheterotrophic plants, large-scale comparative genomics analysis has not been conducted among parasitic or mycoheterotrophic plants or

between these two groups of parasites. Using the high-quality genomes of *G. elata* and *C. australis*, we were able to inspect the convergent loss of genes in these two full heterotrophs with hosts from different kingdoms. We found that in both *G. elata* and *C. australis*, several pathways are commonly non-functional due to gene loss, including photosynthesis, flowering time regulation, immunity, nutrient uptake, and leaf and root development (Figures 5 and 6; Tables S15–S18). These findings provide new insight into the convergent evolution by gene loss in *G. elata* and *C. australis* and raise many new questions, e.g., how *G. elata* regulates its flowering and how it defends against soil pathogens.

Previous studies on the plastomes of various parasitic plants indicated strong contraction in the plastomes of parasitic plants (Feng et al., 2016). Gene family contraction has also been found to be one of the major events of nuclear genome evolution of parasitic plants (Alexandre et al., 2018; Cai et al., 2021; Ohigashi et al., 2019; Sun et al., 2018). We found strong signals of gene family contraction in the orchids: Approximately 75% of the conserved gene families in almost all orchids showed contraction and *G. elata* exhibited the largest extent of gene family contraction (Figure 2). It is very likely that contraction of gene families is associated with the mycoheterotrophic lifestyles in Orchidaceae. Recently, Cai et al. (2021) compared the pattern of gene loss in *Striga*, *Cuscuta*, and *Sapria* and found that an increased level of parasitism is positively correlated with the degree of gene loss. Similarly, our analysis on both mycoheterotrophic and parasitic plants also indicated that an increased degree of heterotrophy is positively associated with the severity of gene loss (Figure 4; Table S8). Gene loss in the hemiparasite *Striga* and the initial mycoheterotroph *Apostasia* are similar (approximately 4–5%) and the complete heterotrophs *G. elata* and *C. australis* both lost approximately 10% of their genes (Figure S5; Table S8). Comparison between the lost genes in *S.*

asiatica and *A. shenzhenica* did not reveal many common genes (17.62 and 15.98%, respectively), suggesting that gene loss in these partial heterotrophs is largely lineage-specific. However, relatively large portions of lost genes in the complete heterotrophs *G. elata* and *C. australis* (35.60 and 42.59%, respectively) are in common (Figure S5; Table S8), and these lost genes are specifically enriched in a number of pathways that are important for leaf and root functions (Figure 4). The dramatic gene loss in the fully mycoheterotrophic *G. elata* and holoparasitic *C. australis* suggest that gene loss in fully heterotrophic plants is perhaps positively selected for – by losing various genes that are normally important for development and growth and resistance or adaption to environmental factors, some of the heterotrophic plants may gain increased fitness. For example, *C. australis* lost many genes important for controlling flowering time (Sun et al., 2018), and by losing its own flowering regulatory machinery, *C. australis* is able to use the host plants' FLOWERING LOCUS T signals to activate its own flowering, so that the flowering times of *C. australis* and hosts are well synchronized (Shen et al., 2020). This unique flowering behavior of *C. australis* is one of the important mechanisms by which *C. australis* and probably many other *Cuscuta* species can parasitize various species of hosts. Notably, even though gene loss in *G. elata* and *C. australis* converges on various pathways, the exact lost genes may not be the same (Figures 5 and 6; Tables S15–S18). A pathway could be defective even if different genes in the pathway are lost or non-functional. Many genes may have multiple functions and are involved in different pathways; for a specific pathway, loss of different genes in *G. elata* and *C. australis* may be a result from the selection pressure of retaining genes that function in other pathways, even if they no longer appear to be functional for the given pathway.

Genome sequencing and comparative analysis suggested that both *G. elata* and *C. australis* are likely not able to synthesize phyloquinone or thiamine, two important vitamins. However, both vitamins were detected in these two fully heterotrophic plants (Figure S9; Table S19). Although *G. elata* seems to have lost its capability of ABA biosynthesis, due to a lack of β -carotene, ABA was detected in *G. elata*. Possibly *G. elata* could transfer ABA from the *Armillaria* host fungi. Similarly, ABA biosynthesis genes were almost completely lost in the endoparasitic *Sapria* (Cai et al., 2021). It would be interesting to determine whether *Sapria* contains ABA; if so, *Sapria* very likely obtains ABA from the host plant as well. Thus, in addition to nutrients and energy, mycoheterotrophic and parasitic plants convergently evolved to rely on hosts for other essential life-sustaining biomolecules, further supporting the scenario that loss of certain genes/pathways and the consequent dependency on hosts could be positively selected for during the evolution of heterotrophic plants.

Obviously mycoheterotrophic and parasitic plants are distinct in their physiology. Yet very little is known about the key genes that are required for establishment and maintenance of the interactions between mycoheterotrophs and host fungi and between parasitic plants and hosts. This has hindered us from analyzing the specific evolution of parasitism genes in both *G. elata* and *C. australis*. Nevertheless, our gene loss analyses indicated several specifically lost pathways in *G. elata* and *C. australis*. For example, *C. australis* lost at least nine key genes that function in symbiosis (Figure 7), but five of them are still retained in *G. elata* (Figure 7; Table S18). This is consistent with the physiology of *C. australis*, which likely no longer needs to interact with soil bacteria or fungi, while Orchidaceae species including *G. elata* are partially or completely dependent on the interaction with its host fungi to survive (Zhang et al., 2018). The endoparasitic *S. himalayana* genome comprises exceptionally long introns (Cai et al., 2021). Compared with *C. australis* and most autotrophic plants, *G. elata* introns are also very large (Figure 3). Although long introns are a feature of most Orchidaceae plants (initial and full mycoheterotrophs) (Figure 3), the basal orchid *A. shenzhenica* and the closely related outgroup *A. officinalis* have a similar feature of relatively short introns. Given that *C. australis* does not have long introns either, it seems that large introns are not a genome feature related to heterotrophy, but a result of lineage-specific evolutionary events.

Almost all orchids are initial mycoheterotrophs, which obtain nutrients from fungi during germination. Thus, initial mycoheterotrophy likely emerged at the ancestral stage of Orchidaceae. We speculate that the initial mycoheterotrophic lifestyle in Orchidaceae may have resulted from neofunctionalization of certain genes after gene family expansion. Although our analysis indicated that the common ancestor of orchids also experienced gene family expansion, functional studies are needed to identify the genes that are important for establishment of initial mycoheterotrophy during orchid seed germination and to infer the evolutionary origin of these critical genes.

In summary, we provide strong evidence to support the scenario that gene loss plays an important role in the evolution of mycoheterotrophic and parasitic plants and that gene loss may not only be a consequence, but also a driving force of the evolution of parasitism. Increasing numbers of genomes of heterotrophic plants and physiological and ecological studies will provide exciting new opportunities for understanding these parasites.

EXPERIMENTAL PROCEDURES

Plant materials

An individual tuber of *G. elata* (Orchidaceae) was originally purchased from the Xiaocaoba marketplace in Zhaotong, Yunnan

Province, China, in 2010, and it was cultivated until seeds were obtained. The seeds were sown into a mixture, composed of the fungi *M. dendrobii* and *Armillaria* and shredded *Quercus dentata* branches, which were used as the substrate for the fungi. In this manner, this specific *G. elata* ecotype was cultivated for three generations. A voucher specimen of *G. elata* was deposited at the Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences (accession No. 1345284). *Cuscuta australis* was obtained from a line maintained by our lab (Sun et al., 2018). Potato (*Solanum lycopersicum*) was purchased from a local market, and *D. catenatum* and *P. equestris* were kindly provided by Dr. Shibao Zhang (Kunming Institute of Botany, CAS).

DNA library preparation and sequencing

A scape from an individual *G. elata* plant (around 30 cm) was harvested and cleaned with tap water and used for DNA extraction. For NGS, DNA libraries were constructed using a VAHTS Universal Plus DNA Library Prep Kit for MGI (Vazyme, Nanjing, China) following the manufacturer's instructions, and the libraries were sequenced on an MGI-SEQ 2000 (MGI Tech, Shenzhen, China) to generate 150-bp paired-end data. In total, 111.36 Gb of data were obtained. A SMRT bell CLR library for DNA long reads was constructed using a SMRT bell Express Template Prep kit 2.0 (Pacific Biosciences, CA), and on a PacBio Sequel II platform 133.33 Gb of subreads were generated, with an average length of 12 313 bp. The same *G. elata* scape tissue was used for the Hi-C pipeline following a previously described method (Rao et al., 2014), including crosslinking, chromatin digestion with *MboI* (New England Biolabs, MA), labeling of DNA ends, ligation, purification, shearing, and biotin pull-down. The Hi-C libraries were sequenced on an MGI-SEQ 2000 (MGI Tech, Shenzhen, China) platform in 150-bp paired-end mode.

Genome survey, assembly, and annotation

We first surveyed the genome of *G. elata* based on *k*-mer depth-frequency distribution analysis using GCE software (v1.0.0) (Liu et al., 2013) with NGS short-read data. Details of the steps and parameters can be found in Methods S1 in the Supporting Information.

In the genome assembly, the subreads from the Pacbio were corrected, trimmed, and assembled by CANU (v1.9) (Koren et al., 2017). Contigs were polished with PacBio data using GCp (v1.9.0) (<https://github.com/PacificBiosciences/gcpp>) and then corrected with NGS data using Pilon (v1.23) (Walker et al., 2014). To obtain chromosome-scale assemblies, we mapped the Hi-C data to contigs and constructed Hi-C contact matrices, which were used to correct the assembly of contigs and build scaffolds using the 3D-DNA pipeline (v180922) (Dudchenko et al., 2017). After obtaining the chromosome-scale assembly, we first annotated the repetitive sequences according to structural features and sequence homology. We then identified the rRNA, tRNA, and other RNA genes in the genome using the structural and sequence features of various RNA genes. Finally, we combined homology annotation, *de novo* annotation, and RNA-seq data-based annotation for the annotation of the coding genes, and finally the protein coding genes were obtained. Complete details are described in Methods S2 and S3 in the Supplemental Information.

Comparative genomics analysis

Gene family expansion and contraction in Orchidaceae, identification of orthologous introns, comparison of gene loss in orchids and *C. australis*, *S. asiatica*, and *S. himalayana*, and identification of Arabidopsis orthologs in *G. elata*, *C. australis*, and their close

relatives are described in Methods S4–S7 in the Supplemental Information.

Real-time quantitative PCR analysis

Real-time qPCR was performed with a CFX Connect Real-Time PCR Detection System (Bio-Rad, USA) by using the qPCR Core kits (Bio-Rad, USA). Gene expression was normalized using *Actin* as a reference gene. The qPCR primers are listed in Table S21.

Detection of phytohormones, vitamins, and carotenoids

Stem segments of *Dendrobium officinale* and *C. australis* and *G. elata* tuber cubes were added into a *P. syringae* pv. tomato DC3000 suspension (OD = 0.5) or water. After 24 h, samples were harvested for SA detection. Phytohormone determination was done following a previously described method (Song et al., 2021). The quantification of vitamins B1 and K1 was performed according to a previously established method (Claussen et al., 2015; Verstraete et al., 2020). Quantification of carotenoids was performed according to a previously established HPLC method (Marinova and Ribarova, 2007).

ACKNOWLEDGMENTS

We thank Dr. Shibao Zhang (Kunming Institute of Botany, Chinese Academy of Sciences [CAS]) for providing *D. catenatum* and *P. equestris* plants and Dr. Quanlei Qiu (Tibet Bomi Gaoyuan Tibetan Gastrodia Industrial Development Co., Ltd.) for providing *Armillaria*. Wei Chang, Mingxia He, and Jianxiang Yang are thanked for technical support. We are grateful to the Service Center for Experimental Biotechnology at the Kunming Institute of Botany, CAS, for supporting plant cultivation. This research was supported by the Strategic Priority Research Program of the CAS (XDPB16), the National Natural Science Foundation of China (32000179), the General and Key Project of the Applied Basic Research Program of Yunnan (202001AS070021), Poverty Alleviation Through Science and Technology Projects of CAS (KFJ-FP-201905), the Technology Transfer into Yunnan Project (202003AD150005), Digitalization, Development, and Application of Biotic Resources (202002AA100007), and the China Postdoctoral Science Foundation (2020M673315).

AUTHOR CONTRIBUTIONS

JW, YX, and YL conceived and designed the research; YX, YL, ZS, and MZ performed experimental work; all authors analyzed the data; YX, YL, and JW wrote the manuscript. All the authors have read the manuscript and have approved its submission.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

The original sequencing data and assembly data of the *G. elata* genome are available at the National Genomics Data Center (<https://ngdc.cnca.ac.cn/>) under BioProject number PRJCA005619.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. *k*-mer ($k = 17$) distribution and genome size estimation by an H0/H1 model using GCE for the *Gastrodia elata* genome.

Figure S2. Identification of contamination and organelle sequences using the GC ratio, the genome survey sequence data mapping depth, and the results of organellar genome BLAST searches.

Figure S3. Hi-C contact map of *Gastrodia elata*.

Figure S4. Comparison of gene annotation from this study with that from Yuan *et al.*

Figure S5. Venn diagrams of lost genes in three myco-heterotrophic plants and three parasitic plants.

Figure S6. The maximum efficiency of photosystem II photochemistry (Fv/Fm) in *Gastrodia elata*, *Dendrobium officinale*, and *Arabidopsis*.

Figure S7. Levels of salicylic acid in *Gastrodia elata*, *Phalaenopsis equestris*, and *Dendrobium catenatum* after being challenged by a bacterial pathogen.

Figure S8. Gene loss in the thiamine (vitamin B1) and phyloquinone (vitamin K1) biosynthesis pathways.

Figure S9. Gene loss in the MEP and mevalonic acid pathways.

Table S1. Statistics of contig-level assembly length, accuracy, and heterozygosity of the *Gastrodia elata* genome.

Table S2. Transcriptome data for assessment of genome assembly completeness and gene annotation.

Table S3. Protein-coding genes, RNA genes, and repetitive elements in the *Gastrodia elata* genome.

Table S4. Results of BUSCO analysis for the genome and protein-coding gene set of *Gastrodia elata* from this work and Yuan *et al.*

Table S5. Plant genomes for comparative genomics analysis.

Table S6. Gene family expansion and contraction.

Table S7. Statistic information of introns.

Table S8. Gene loss in mycoheterotrophic and parasitic plants.

Table S9. *Arabidopsis* orthologs in *Gastrodia elata*, *Cuscuta australis*, and the reference species.

Table S10. GO enrichment of the orthogroups whose orthologous members are conserved in autotrophs, but lost in both *Gastrodia elata* and *Cuscuta australis*.

Table S11. MapMan annotation of all genes in mycoheterotrophic plants, parasitic plants, and the reference species.

Table S12. Presence/absence of plastid tRNA, rRNA, and protein-coding genes in *Arabidopsis thaliana*, *Oryza sativa*, *Asparagus officinalis*, *Vanilla planifolia*, *Striga hermonthica*, *Cuscuta australis*, and *Gastrodia elata*.

Table S13. Presence and absence of genes involved in photosynthesis.

Table S14. Presence and absence of genes involved in the regulation of flowering time.

Table S15. Presence and absence of genes involved in the salicylic acid pathway.

Table S16. Presence and absence of genes involved in nutrient uptake and transport.

Table S17. Presence and absence of genes involved in leaf and root development.

Table S18. Presence and absence of genes involved in plant symbiosis signaling pathways.

Table S19. Presence and absence of genes involved in thiamine (vitamin B1) and phyloquinone (vitamin K1) biosynthesis.

Table S20. Presence and absence of genes involved in the MEP and mevalonic acid pathways.

Table S21. The list of qPCR primers.

Methods S1. Genome survey.

Methods S2. Genome assembly.

Methods S3. Annotation of genome features.

Methods S4. Gene family expansion and contraction.

Methods S5. Identification of orthologous introns.

Methods S6. Comparison of gene loss in orchids and *C. australis*, *S. asiatica*, and *S. himalayana*.

Methods S7. Identification of *Arabidopsis* orthologs in *G. elata*, *C. australis*, and their reference plants.

REFERENCES

- Alexandre, N.M., Humphrey, P.T., Gloss, A.D., Lee, J., Frazier, J., Affeldt, H.A. *et al.* (2018) Habitat preference of an herbivore shapes the habitat distribution of its host plant. *Ecosphere*, **9**, e02372. <https://doi.org/10.1002/ecs2.2372>.
- Basset, G.J., Latimer, S., Fatihi, A., Soubeyrand, E. & Block, A. (2017) Phyloquinone (Vitamin K1): occurrence, biosynthesis and functions. *Mini-Reviews in Medicinal Chemistry*, **17**, 1028–1038. <https://doi.org/10.2174/1389557516666160623082714>.
- Beaudoin, G.A.W., Johnson, T.S. & Hanson, A.D. (2018) The PLUTO plastidial nucleobase transporter also transports the thiamin precursor hydroxymethylpyrimidine. *Bioscience Reports*, **38**, BSR20180048. <https://doi.org/10.1042/BSR20180048>.
- Blumel, M., Dally, N. & Jung, C. (2015) Flowering time regulation in crops—what did we learn from *Arabidopsis*? *Current Opinion in Biotechnology*, **32**, 121–129. <https://doi.org/10.1016/j.copbio.2014.11.023>.
- Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W. *et al.* (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nature Genetics*, **47**, 65–72. <https://doi.org/10.1038/ng.3149>.
- Cai, L., Arnold, B.J., Xi, Z., Khost, D.E., Patel, N., Hartmann, C.B. *et al.* (2021) Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). *Current Biology*, **31**, 1002–1011. <https://doi.org/10.1016/j.cub.2020.12.045>.
- Chase, M.W., Cameron, K.M., Freudenstein, J.V., Pridgeon, A.M., Salazar, G., van den Berg, C. *et al.* (2015) An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society*, **177**, 151–174. <https://doi.org/10.1111/boj.12234>.
- Chen, S., Wang, X., Wang, Y., Zhang, G., Song, W., Dong, X. *et al.* (2020) Improved *de novo* assembly of the achlorophyllous Orchid *Gastrodia elata*. *Frontiers in Genetics*, **11**, 580568. <https://doi.org/10.3389/fgene.2020.580568>.
- Clarke, C.R., Timko, M.P., Yoder, J.I., Axtell, M.J. & Westwood, J.H. (2019) Molecular dialog between parasitic plants and their hosts. *Annual Review of Phytopathology*, **11**, 1–21. <https://doi.org/10.1146/annurev-phyto-082718-100043>.
- Claussen, F.A., Taylor, M.L., Breeze, M.L. & Liu, K. (2015) Measurement of vitamin K1 in commercial canola cultivars from growing locations in north and south america using High-Performance Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, **63**, 1076–1081. <https://doi.org/10.1021/jf503824t>.
- Ding, P. & Ding, Y. (2020) Stories of salicylic acid: A plant defense hormone. *Trends in Plant Science*, **25**, 549–565. <https://doi.org/10.1016/j.tplants.2020.01.004>.
- Du, F., Guan, C. & Jiao, Y. (2018) Molecular mechanisms of leaf morphogenesis. *Molecular Plant*, **11**, 1117–1134. <https://doi.org/10.1016/j.molp.2018.06.006>.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C. *et al.* (2017) *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**, 92–95. <https://doi.org/10.1126/science.aal3327>.
- Estevez, J.M., Cantero, A., Reindl, A., Reichler, S. & Leon, P. (2001) 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *Journal of Biological Chemistry*, **276**, 22901–22909. <https://doi.org/10.1074/jbc.M100854200>.
- Feng, Y.-L., Wicke, S., Li, J.-W., Han, Y.U., Lin, C.-S., Li, D.-Z. *et al.* (2016) Lineage-specific reductions of plastid genomes in an orchid tribe with

- partially and fully mycoheterotrophic species. *Genome Biology and Evolution*, **8**, 2164–2175. <https://doi.org/10.1093/gbe/evw144>.
- Greenwood, M. & Locke, J.C. (2020) The circadian clock coordinates plant development through specificity at the tissue and cellular level. *Current Opinion in Plant Biology*, **53**, 65–72. <https://doi.org/10.1016/j.pbi.2019.09.004>.
- Han, S.-K., Qi, X., Sugihara, K., Dang, J.H., Endo, T.A., Miller, K.L. et al. (2018) MUTE directly orchestrates cell-state switch and the single symmetric division to create stomata. *Developmental Cell*, **45**(303–315), e305. <https://doi.org/10.1016/j.devcel.2018.04.010>.
- Hasing, T., Tang, H., Brym, M., Khazi, F., Huang, T. & Chambers, A.H. (2020) A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nature Food*, **1**, 811–819. <https://doi.org/10.1038/s43016-020-00197-2>.
- Hu, Z., Yang, Z. & Wang, J. (1988) Isolation and partial characterization of an antifungal protein from *Gastrodia elata* corm. *Acta Botanica Yunnanica*, 373–380.
- Huisman, R., Hontelez, J., Mysore, K.S., Wen, J., Bisseling, T. & Limpens, E. (2016) A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform controls the formation of a stable host-microbe interface in symbiosis. *New Phytologist*, **211**, 1338–1351. <https://doi.org/10.1111/nph.13973>.
- Jiang, Y., Wang, W., Xie, Q., Liu, N.A., Liu, L., Wang, D. et al. (2017) Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. *Science*, **356**, 1172–1175. <https://doi.org/10.1126/science.aam9970>.
- Kong, D., Zhu, Y., Wu, H., Cheng, X., Liang, H. & Ling, H.Q. (2008) *AtTHIC*, a gene involved in thiamine biosynthesis in *Arabidopsis thaliana*. *Cell Research*, **18**, 566–576. <https://doi.org/10.1038/cr.2008.35>.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. & Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, **27**, 722–736. <https://doi.org/10.1101/gr.215087.116>.
- Leake, J.R. (1994) The biology of myco-heterotrophic (saprophytic) plants. *New Phytologist*, **127**, 171–216. <https://doi.org/10.1111/j.1469-8137.1994.tb04272.x>.
- Leake, J.R. (2004) Myco-heterotroph/epiparasitic plant interactions with ectomycorrhizal and arbuscular mycorrhizal fungi. *Current Opinion in Plant Biology*, **7**, 422–428. <https://doi.org/10.1016/j.pbi.2004.04.004>.
- Li, Y.-X., Li, Z.-H., Schuiteman, A., Chase, M.W., Li, J.-W., Huang, W.-C. et al. (2019) Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **139**, 106540. <https://doi.org/10.1016/j.ympev.2019.106540>.
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H. & Li, N. et al. (2013) Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects.
- Liu, C.-W., Breakspear, A., Stacey, N., Findlay, K., Nakashima, J., Ramakrishnan, K. et al. (2019) A protein complex required for polar growth of rhizobial infection threads. *Nature Communications*, **10**, 2848. <https://doi.org/10.1038/s41467-019-10029-y>.
- Luginbuehl, L.H., Menard, G.N., Kurup, S., Van Erp, H., Radhakrishnan, G.V., Breakspear, A. et al. (2017) Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. *Science*, **356**, 1175–1178. <https://doi.org/10.1126/science.aan0081>.
- Machado, C.R., de Oliveira, R.L., Boiteux, S., Praekelt, U.M., Meacock, P.A. & Menck, C.F. (1996) *Thi1*, a thiamine biosynthetic gene in *Arabidopsis thaliana*, complements bacterial defects in DNA repair. *Plant Molecular Biology*, **31**, 585–593. <https://doi.org/10.1007/BF00042231>.
- Marinova, D. & Ribarova, F. (2007) HPLC determination of carotenoids in Bulgarian berries. *Journal of Food Composition and Analysis*, **20**, 370–374. <https://doi.org/10.1016/j.jfca.2006.09.007>.
- McCormick, M.K., Whigham, D.F. & Canchani-Viruet, A. (2018) Mycorrhizal fungi affect orchid distribution and population dynamics. *New Phytologist*, **219**, 1207–1215. <https://doi.org/10.1111/nph.15223>.
- Merckx, V.S.F.T., Janssens, S.B., Hynson, N.A., Specht, C.D., Bruns, T.D. & Smets, E.F. (2012) Mycoheterotrophic interactions are not limited to a narrow phylogenetic range of arbuscular mycorrhizal fungi. *Molecular Ecology*, **21**, 1524–1532. <https://doi.org/10.1111/j.1365-294X.2012.05472.x>.
- Motte, H., Vanneste, S. & Beekman, T. (2019) Molecular and environmental regulation of root development. *Annual Review of Plant Biology*, **70**, 465–488. <https://doi.org/10.1146/annurev-arplant-050718-100423>.
- Murray, J.D., Muni, R.R.D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M. et al. (2011) *Vapyrin*, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. *Plant Journal*, **65**, 244–252. <https://doi.org/10.1111/j.1365-3113X.2010.04415.x>.
- Nambara, E. & Marion-Poll, A. (2005) Abscisic acid biosynthesis and catabolism. *Annual Review of Plant Biology*, **56**, 165–185. <https://doi.org/10.1146/annurev-arplant.56.032604.144046>.
- Nisar, N., Li, L., Lu, S., Khin, N.C. & Pogson, B.J. (2015) Carotenoid metabolism in plants. *Molecular Plant*, **8**, 68–82. <https://doi.org/10.1016/j.molp.2014.12.007>.
- Noordally, Z.B., Trichtinger, C., Dalvit, I., Hofmann, M., Roux, C., Zamboni, N. et al. (2020) The coenzyme thiamine diphosphate displays a daily rhythm in the *Arabidopsis* nucleus. *Communications Biology*, **3**, 209. <https://doi.org/10.1038/s42003-020-0927-z>.
- Ohgashi, K., Mizuguti, A., Nakatani, K., Yoshimura, Y. & Matsuo, K. (2019) Modeling the flowering sensitivity of five accessions of wild soybean (*Glycine soja*) to temperature and photoperiod, and its latitudinal cline. *Breed Science*, **69**, 84–93. <https://doi.org/10.1270/jsbbs.15-136P>.
- Oldroyd, G.E. (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews: Microbiology*, **11**, 252–263. <https://doi.org/10.1038/nrmicro2990>.
- Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J. et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>.
- Sanchez, S.E., Rugnone, M.L. & Kay, S.A. (2020) Light perception: a matter of time. *Molecular Plant*, **13**, 363–385. <https://doi.org/10.1016/j.molp.2020.02.006>.
- Shen, G., Liu, N., Zhang, J., Xu, Y., Baldwin, I.T. & Wu, J. (2020) *Cuscuta australis* (dodder) parasite eavesdrops on the host plants' FT signals to flower. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 23125–23130. <https://doi.org/10.1073/pnas.2009445117>.
- Shimaoka, C., Fukunaga, H., Inagaki, S. & Sawa, S. (2017) Artificial cultivation system for *Gastrodia* spp. and identification of associated mycorrhizal fungi. *International Journal of Biology*, **9**, 27–34. <https://doi.org/10.5539/ijb.v9n4p27>.
- Smith, J.D., Johnson, B.I., Mescher, M.C. & De Moraes, C.M. (2021) A plant parasite uses light cues to detect differences in host-plant proximity and architecture. *Plant, Cell and Environment*, **44**, 1142–1150. <https://doi.org/10.1111/pce.13967>.
- Song, J., Bian, J., Xue, N., Xu, Y. & Wu, J. (2021) Inter-species mRNA transfer among green peach aphids, dodder parasites, and cucumber host plants. *Plant Diversity*, <https://doi.org/10.1016/j.pld.2021.03.004>.
- Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C. et al. (2018) Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nature Communications*, **9**, 2683. <https://doi.org/10.1038/s41467-018-04721-8>.
- Verstraete, J., Strobbe, S., Van Der Straeten, D. & Stove, C. (2020) The first comprehensive LC-MS/MS method allowing dissection of the thiamine pathway in plants. *Analytical Chemistry*, **92**, 4073–4081. <https://doi.org/10.1021/acs.analchem.9b05717>.
- Vogel, A., Schwacke, R., Denton, A.K., Usadel, B., Hollmann, J., Fischer, K. et al. (2018) Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nature Communications*, **9**, 2515. <https://doi.org/10.1038/s41467-018-04344-z>.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**, e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Wang, X., Bauw, G., Van Damme, E.J.M., Peumans, W.J., Chen, Z.-L., Van Montagu, M. et al. (2001) Gastrodin-like mannose-binding proteins: a novel class of plant proteins with antifungal properties. *Plant Journal*, **25**, 651–661. <https://doi.org/10.1046/j.1365-3113X.2001.00999.x>.
- Wang, Y., Chen, Y.F. & Wu, W.H. (2021) Potassium and phosphorus transport and signaling in plants. *Journal of Integrative Plant Biology*, **63**, 34–52. <https://doi.org/10.1111/jipb.13053>.
- Wang, Y.Y., Cheng, Y.H., Chen, K.E. & Tsay, Y.F. (2018) Nitrate transport, signaling, and use efficiency. *Annual Review of Plant Biology*, **69**, 85–122. <https://doi.org/10.1146/annurev-arplant-042817-040056>.

- Wang, Z., Ruan, W., Shi, J., Zhang, L., Xiang, D., Yang, C. *et al.* (2014) Rice SPX1 and SPX2 inhibit phosphate starvation responses through interacting with PHR2 in a phosphate-dependent manner. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 14953–14958. <https://doi.org/10.1073/pnas.1404680111>.
- Westwood, J.H., Yoder, J.I., Timko, M.P. & dePamphilis, C.W. (2010) The evolution of parasitism in plants. *Trends in Plant Science*, **15**, 227–235. <https://doi.org/10.1016/j.tplants.2010.01.004>.
- Xu, J., Ran, Y. & Guo, S. (1989) Studies on the life cycle of *Gastrodia elata*. *Acta Academiae Medicinae Sinicae*, **11**, 237–241+322–325.
- Yoshida, S., Kim, S., Wafula, E.K., Tanskanen, J., Kim, Y.-M., Honaas, L. *et al.* (2019) Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Current Biology*, **29**, 3041–3052. <https://doi.org/10.1016/j.cub.2019.07.086>.
- Yu, S., Cao, L.I., Zhou, C.-M., Zhang, T.-Q., Lian, H., Sun, Y. *et al.* (2013) Sugar is an endogenous cue for juvenile-to-adult phase transition in plants. *Elife*, **2**, e00269. <https://doi.org/10.7554/eLife.00269>.
- Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X. *et al.* (2018) The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nature Communications*, **9**, 1615. <https://doi.org/10.1038/s41467-018-03423-5>.
- Zhan, H.-D., Zhou, H.-Y., Sui, Y.-P., Du, X.-L., Wang, W.-H., Dai, L.I. *et al.* (2016) The rhizome of *Gastrodia elata* Blume - An ethnopharmacological review. *Journal of Ethnopharmacology*, **189**, 361–385. <https://doi.org/10.1016/j.jep.2016.06.057>.
- Zhang, G.-Q., Liu, K.-W., Li, Z., Lohaus, R., Hsiao, Y.-Y., Niu, S.-C. *et al.* (2017) The *Apostasia* genome and the evolution of orchids. *Nature*, **549**, 379–383. <https://doi.org/10.1038/nature23897>.
- Zhang, G.-Q., Xu, Q., Bian, C., Tsai, W.-C., Yeh, C.-M., Liu, K.-W. *et al.* (2016) The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific Reports*, **6**, 19029. <https://doi.org/10.1038/srep19029>.
- Zhang, S., Yang, Y., Li, J., Qin, J., Zhang, W., Huang, W. *et al.* (2018) *Physiological Diversity of Orchids*. *Plant Diversity*, **40**, 196–208. <https://doi.org/10.1016/j.pld.2018.06.003>.
- Zhou, X. (1981) The life cycle of *Gastrodia elata* Blume. *Acta Botanica Yunnanica*, **3**, 197–202.